Table of Contents

United States
Department of
Agriculture

Food and
Nutrition
Service

Office of
Analysis and
Evaluation

THE FOOD STAMP PROGRAM

QUALITY CONTROL SYSTEM

A Report to the U.S. Congress, May 1987

Table of Contents

ACKNOWLEDGEMENTS

This is a comprehensive report that reflects the efforts of many people. Christine Schmidt, Director, Analysis Staff, provided overall direction to the project and edited the report. Abigail Nichols, Chief, Family Nutrition Analysis Branch coordinated research and writing efforts. Alan Hershey of Mathematica Policy Research, Inc., was the lead writer for the report. Steven Carlson directed the evaluation of the statistical aspects of the Food Stamp Program quality control system and was a substantial contributor to planning and editing this report. Steven R. Cole was the FNS project officer for contracted substudies by Abt Associates, Inc., and Mathematica Policy Research, Inc., that were incorporated into this report. John Lee, FNS, was research assistant for the study. Christine Kissmer, Robert Dalrymple, Mark Johnston, Theodore Macaluso also contributed to the study at critical junctures.

The final report reflects substantial assistance and reviews provided by the FNS regions and by the Program Development and Program Accountability Divisions of FNS, particularly Bonny O'Neil, Claire Lipsman, Joseph Pinto, Duane Maddox, Gary Batko and Maurice Tracy. Mack Storrs, QC Study Director for the Department of Health and Human Services (DHHS) acted as study liaison at DHHS. William Hamilton, Abt Associates Inc., advised researchers throughout the study. The names of the many staff of contract research firms appear in the body of the report in references to the substudies upon which this report was constructed. Credit for the final production of this report goes to Lena Cunningham, Sheana Thomas, and Sharon Corbin of Mathematica Policy Research, Inc.

CONTENTS

Table of Contents

LIST OF FIGURES

Figure

Table of Contents

LIST OF TABLES

EXECUTIVE SUMMARY

Study Mandate

This study reponds to a Congressional request for an examination of quality control in the Food Stamp Program. In the Food Security Act of 1985, Congress instructed the U.S. Department of Agriculture (USDA) to undertake two independent studies of the food stamp quality control system: one to be conducted by the Department itself, and one to be conducted by the National Academy of Sciences. The Congress asked that both studies address how the QC system can best be operated, so as (1) to obtain information that allows the State agencies to improve the quality of their program administration, and (2) to provide reasonable data that form the basis for which federal funding may be withheld from State agencies which exhibit excessive levels of erroneous payments. This report presents the results of the Department of Agriculture's study.

Background

The Food Stamp Program is a nationwide program which helps low-income families and individuals buy the food they need to maintain a nutritious diet. The program is authorized by Congress, administered nationally by the U.S. Department of Agriculture's Food and Nutrition Service (FNS), and run through State welfare agencies and their local offices.

In the broadest sense, Congress, through its legislative actions, defines the policy objectives that govern how the Food Stamp Program is administered. FNS translates those policies into regulatory requirements, and the States operate the program, with final responsibility for the day-to-day activities of determining eligibility and benefits. The federal government provides most of the funding for the program, paying the full costs of benefits and about half of the administrative costs of the States and local offices.

During Fiscal Year 1985 alone, the Food Stamp Program dispensed nearly $11 billion in food stamp benefits to roughly 35 million individuals. With so much money involved, even small error rates can have very large costs. Thus, with an overpayment error-rate of just over 8 percent in 1985, nearly $900 million was issued above and beyond the amount that households should have received under law. At the same time, with an underpayment-error rate of just over 2 percent, States should have issued another $250 million to households that were already determined eligible.

Food Stamp Program Accountability

To realize the basic objectives of and to ensure accountability in the program, FNS set up an array of performance monitoring systems. Management Evaluation (ME) reviews, for example, monitor the compliance of States and local offices with a wide range of administrative requirements (such as the timeliness of certification and issuance actions and the accessibility of office locations and the hours of operation). A set of audit and accounting procedures monitor administrative costs. Moreover, both the General Accounting Office--the non-partisan investigative arm of Congress--and USDA's Office of the Inspector General conduct routine and special audits of program operations. As required by law, the food stamp quality control (QC) system monitors the accuracy of eligibility and benefit determinations.

The food stamp quality control system, as it stands today, reflects more than 15 years of development, legislative oversight, and refinement. Major milestones in its development include the publication of the first regulations in 1971; the implementation of the comprehensive Performance Reporting System (PRS) in 1975; specific legislative authorization for the PRS and QC in the Food Stamp Act of 1977; and the introduction of an error rate liability and incentive system in the legislative amendments of 1980 and 1982.

The Food Stamp Quality Control System

The Food Stamp Program's quality control system provides two general measures of certification accuracy. The first is based on an intensive review of a sample of program participants. Each review determines whether the household is eligible and receiving the correct food stamp benefit--neither more nor less--given its income, expenses, resources, and living arrangements. The second measure of certification accuracy is based on a sample of households whose application for food stamps is denied or whose benefits are terminated. This negative action review determines whether the decision to deny or terminate benefits is procedurally correct and fully documented.

Quality control reviews provide a basis for a system of financial liabilities and incentives whose purpose is to hold States accountable for certification accuracy. Under current law, States whose overpayment error rates are beyond a specified threshold are liable for a portion of the cost of the overpayments. States that exhibit low error rates (including underpayments and negative actions) are eligible for additional federal funding. Liabilities and incentives provide a mechanism by which both State and federal governments share in the cost of certification errors.

Because food stamp benefits are fully funded by federal tax dollars, the federal government would bear the full cost of all erroneous payments in the absence of quality control liabilities. Thus, the liabilities limit the fiscal burden imposed on the federal government by redistributing some of the risk of erroneous certification decisions to State agencies. Even under the current system, the federal government absorbs the full cost of overpayment up to 5 percent of all benefits (roughly $500 million a year). States were liable for less than 25 percent of the cost of overpayments in 1985.

Scope of the Report

State concerns about the reliability and equity of the QC process have increased in recent years as both the number of States liable for overpayments and the amount of their liabilities have grown.2/ The Congressional directive that FNS and the National Academy of Sciences evaluate the current quality control system reflects these concerns.

The scope of this study is defined by the Congressional mandate. The mandate is explicit in its direction to study QC as the basis for withholding funds from States that exhibit excessive error rates. While there is room for debate about the definition of error, the procedures for measuring errors, and the manner in which liabilities are assessed, it is clear that Congress views QC liabilities as a vehicle for limiting the fiscal participation of the federal government in erroneous payments. The study, therefore, does not deal with broader measures of program performance beyond certification accuracy. While other program objectives are important in their own right, none is intrinsic to the objectives and operations of QC as defined by Congress.

Given this emphasis on certification accuracy, the study examines two applications of quality control: performance measurement and management information. In many respects, the two are closely linked. The information produced by the QC system, if properly classified, presented, and fed back to program managers, can be an important factor in improving certification accuracy. The more closely that performance measurement and management information are integrated, the more effective such feedback can be. In other respects, however, the requirements of performance measurement and management information may make somewhat different demands on

---

[2]In 1981, 16 States were liable for $29 million. By 1985, 48 States were liable for over $200 million.

the design of a quality control system. In weighing these tradeoffs, it is important to remember that QC is the only comprehensive source of performance measures for certification accuracy, while many sources of management information are available.

The specific issues that have been examined fall into two general categories: largely technical issues for which empirical analysis can inform the decision process, and policy issues for which there are no strictly technical answers. The first set includes such issues as the statistical properties of the estimated error rates, and the effects of the caseload and local area on error rates.

The second class of issues is less easily illuminated by empirical analysis and cannot be resolved with strictly technical answers. These issues include questions about the degree of sampling error which should be accepted, the extent to which State and federal governments should share in the cost of certification errors, and the manner in which estimates of error rates should be used to allocate the cost of errors between States and the federal government.

Key Findings

The design of food stamp quality control is fundamentally sound.

- The procedures for drawing samples are standard and widely accepted sampling methods.

- The procedures for estimating error rates are statistically valid. Extensive simulations--the equivalent of 1,000 years of drawing QC samples and estimating error rates--demonstrate that the estimates of overpayment-error rates are essentially unbiased estimates of the true error rate.

- Estimated error rates will vary from sample to sample within a State simply because of sampling errors. In 1985, the standard errors of the estimated overpayment-error rate--a measure of sampling variability--ranged from .004 to .032, with an average of about .01. What constitutes an "acceptable" level of sampling error depends on the cost of obtaining more precise estimates, how the estimates are used, and the consequences to which they lead.

The implementation of QC is generally accurate, reliable, and consistent.

- Repeated audits by the General Accounting Office and the USDA Office of the Inspector General indicate a generally high level of conformance in the application of QC policy by State and federal reviewers. The Food and Nutrition Service provides technical assistance and other efforts to ensure the necessary consistency.

- An analysis of State QC expenditures and reported error rates shows no significant correlation, and thus no evidence that differences in the "level of effort" applied to the review process lead to systematic differences in error rates.

The definition of error is reasonably complete and balanced.

- Strictly procedural or technical errors--those without financial consequences--are excluded from the food stamp payment-error rate. This exclusion is consistent with the focus of QC on the outcome--rather than the process--of the eligibility and benefit determination.

- All other errors with financial consequences--whether caused by the State agency or the participant--are included in the food stamp payment-error rate. The distinction between an agency error and a client error is often ambiguous, and subject to substantial judgment and interpretation. A blanket exclusion of all "client-caused" errors would ignore the variety of actions that can be and are taken by State agencies to reduce the rate of such errors.

- Available evidence does not indicate that emphasis on reducing overpayments leads to an increase in underpayments. The inclusion of a measure of underpayment errors in the calculation of liabilities, however, could alleviate some concerns about the potentially adverse consequences of the current legislative emphasis on overpayment errors.

- While conceptually appealing, the broadest measure of underpayments--one that incorporates the "cost" of an improper denial or termination of households that are truly eligible for benefits--is not feasible under the current QC design. The current review of negative actions is strictly procedural, measuring compliance with process and documentary requirements. Major modifications to the current negative action review process would be necessary to measure the true "cost" of improper decisions.

Certain caseload and area characteristics do affect the incidence and size of errors (with some caveats). There is no clear basis for adjusting State error rates (or, equivalently, State error thresholds) for differences in the complexity of the caseload or the characteristics of the local area. No statistical model yet identified can adjust State error rates and produce better measures of error.

- States vary in a number of ways, some of which lead to higher error rates, while others lead to lower rates. Errors are affected by a complex interaction among variables that offset or enhance each other to varying degrees across and within States. This means that adjusting error rates by a few indicators with prespecified adjustments will not improve the measure of a State's error rate.

- An extensive multivariate analysis reveals that household size, the presence and source of income and assets, the number of deductions, and the density of the population around the local office were associated with the size of the average food stamp benefit, the probability of error, or the size of the error. Some important variables, particularly measures of caseload dynamics, could not be included in the analysis. In their absence, these results must be interpreted with some caution.

- The analysis considers several different models whose usefulness is similar in a purely statistical sense. Selecting the single, most appropriate adjustment model from among those presented is, however, problematic. Individual States fare differently under each model, and small changes in adjusted error rates lead to large changes in liability under current methods for assessing liability.

The legislative basis for calculating liabilities could be improved.

- There is no empirical basis for determining the "right" threshold. Setting the error rate target is a policy decision that must balance the tradeoffs between the desire for program accuracy and the recognition of some level of unavoidable error.

- An analysis of the differences between AFDC and the historically higher food stamp error rates suggests no reason that the target should be the same in both programs.

Differences in the caseload help explain most of the difference in error rates.

- Using administrative costs as the basis for liabilities could discourage administrative investment. Liabilities are set as a proportion of the federal share of administrative costs. Thus, if two States exhibit the same error rate, the one with higher administrative costs will have a larger liability. A benefit basis for liabilities could link the liability more closely to the actual cost of the erroneous payments and eliminate this potential disincentive.

- With the current legislative formula to calculate liabilities, trivial differences in error rates can make important differences in the size of the liability. Conversely, some differences in error rates are ignored. Moreover, because of the sampling variability inherent in the estimates of error rates, the step function systematically overstates liabilities for some States and understates liabilities for others. A continuous, smooth function could eliminate these features of the current system.

- The point estimate reflects the best available measure of a State's performance and is, therefore, the best measure on which to base liabilities. A lower confidence bound systematically and substantially understates the true liability (i.e., the liability that would be assessed if the true overpayment-error rate were known). Using the lower bound would not improve equity. Instead, it would simply increase the risks to the Federal government.

- Using the point estimate will sometimes produce liabilities that are too high or too low for a State in any one time period, because of sampling error. Under current procedures, the federal government bears the full cost of the first 5 percent of overpayments. Beyond that point, the risks are shared by the State and federal governments. The risk to the State increases as true error rates approach the threshold.

The current QC system lays the foundation for a relatively comprehensive, accuracy-oriented management information system for State managers. To some extent, relatively small State sample sizes and limited data on administrative procedures restrict the usefulness of the current QC system for management information.

- Most States use QC for analytical purposes or as a vehicle for collecting additional management information.

- Current sample sizes limit the usefulness of QC for analyzing local office error rates or for obtaining interim measures of progress. However, many areas accumulate sample results over time, thus increasing the effective sample size for analytic purposes. States can also increase sample sizes, and the federal government pays half the cost.

- QC data provide reasonably complete information on the characteristics of the caseload and types of error, but more limited information on procedures that cause or prevent errors.

Many of the issues raised in this study depend on policy choices. On some issues, the weight of the evidence enables USDA to reach clear conclusions. In other areas, the appropriate approaches are less obvious, and more than one alternative may be acceptable. Pending a thorough review of the parallel report prepared by the National Academy of Sciences, the Department has elected not to make specific recommendations.

I. INTRODUCTION

This study responds to a Congressional request for an examination of quality control in the Food Stamp Program. In the Food Security Act of 1985 (P.L. 99-198), the 99th Congress instructed the U.S. Department of Agriculture (USDA) to undertake two independent studies of the food stamp quality control system--one to be conducted by the Department itself, and one study to be conducted under contract by the National Academy of Sciences.1/ The Congress asked that both studies address

[redacted]

information that will allow the State agencies to improve the quality of program administration, and (2) to provide reasonable data that will form the basis for determining whether federal funding should be withheld for State agencies that exhibit excessive levels of erroneous payments. This report presents the results of the Department of Agriculture's study.

In general, the fundamental aim of quality control is to prevent chronic problems in the production of goods or the delivery of services by measuring performance, comparing performance with established standards, and pursuing corrective action where necessary to improve performance. The concept of quality control has its modern roots in the development of statistical theory in the 1920s. Its use spread through private industry and government operations during World War II in response to the demands of the war effort for high-quality mass production.

The Food Stamp Program (FSP) adopted quality control (QC) in 1971, and has been refining it since then. QC began as an internal management initiative, but its subsequent development involved a series of Congressional actions and the associated evolution of regulations. From the beginning, there has been continued analysis of the success and implications of QC from the standpoint of both federal and State agencies, and technical evaluations by government agencies and independent consultants. Throughout its history, the FSP quality control

the public's confidence in the integrity of the Food Stamp Program.

The importance attached to the food stamp quality control process from its origins in 1971 reflects three key features of the program:

1. Program Size. During Fiscal Year 1985, the Food Stamp Program issued nearly $11 billion in benefits to nearly 35 million recipients. With so much money involved, even small error rates can have large costs. And with so many persons affected and so many decisions that affect them, a method for preventing errors is essential to avoid issuing or denying benefits incorrectly. Specifically, with an overpayment error rate just over 8 percent in 1985, nearly $900 million was issued in excess of the amount that households should have received under the law; at the same time, States should have issued another $250 million to households already determined eligible.

2. Program Growth. The program's rapid growth from a pilot project in the early 1960s to a major national program in the 1970s contributed to concerns about the program's ability to ensure that benefits are issued accurately and that benefits are made available to every eligible household that applies. Results from early QC reviews suggested that overpayments equaled more than 15 percent of total benefits issued, and that underpayments equaled about 3 percent.2/

3. Fiscal Responsibilities. The Food Stamp Program is unique among the major means-tested transfer programs in that administrative responsibilities are shared between the federal and State governments, but benefits are paid for entirely by federal funds. In other transfer programs, State responsibility for funding at least a portion of the benefits that are issued creates natural incentives for quality control and error prevention. As in any delegation of authority, demands arise for program monitoring and accountability mechanisms by the funding agency.

2/These early results pertained to non-public assistance households (i.e., those receiving neither Aid to Families with Dependent Children nor Supplemental Security Income).

The food stamp quality control system examined in this report is the result of over 15 years of development in response to these concerns, and reflects both technical development and difficult policy choices. The remainder of this introductory chapter provides the background for the report that follows. Section A highlights the present functions and purposes of food stamp quality control. Section B defines the questions to be examined in this report, summarizing the questions posed by Congress and describing the process by which FNS developed its study agenda. Section C describes the organization of the remainder of the report.

A. FUNCTIONS AND PURPOSES OF QUALITY CONTROL

Quality control is an integral piece of the broader Performance Reporting System for the Food Stamp Program. The Performance Reporting System was implemented by regulation in 1975 and then mandated by Congress as part of a far-reaching set of changes to the program made in the Food Stamp Act of 1977. The Performance Reporting System includes regular reviews of the program's success in meeting a broader set of program goals: the timeliness of benefit issuance, the accessibility of the program to applicants, and the accuracy of benefit issuance. Separate processes for reviewing and measuring performance have been developed to assess success according to these different dimensions.3/

The Food Stamp Program QC system focuses on the accuracy of household eligibility and benefit determinations, or certifications.4/ It provides two general measures of the accuracy of certification. The first is based on an intensive review of a probability sample of Food Stamp Program participants. These reviews determine whether the participating household is eligible and receiving the correct food stamp benefit--neither more nor less--given its income, expenses, resources, and

---

3/Furthermore, some of these processes go beyond the review and measurement stage. The Performance Reporting system permits financial assessments if States fail to correct performance deficiencies.

4/The process of accepting applications, determining eligibility and benefits, processing interim changes in household circumstances, and periodically reviewing these decisions is known generically as "certification." Throughout this report, "certification accuracy" and "payment accuracy" are used interchangeably.

living arrangements. The second measure is based on a sample of households whose application for food stamps has been denied or whose benefits have been terminated. These reviews of "negative actions" determine whether the decision to deny or terminate benefits was procedurally correct and fully documented.5/

Since the inception of quality control in the Food Stamp Program, the information generated by the reviews of food stamp households and the certification errors associated with them has served two functions. First, it generates objective measures of the administrative performance of States--error rates--that can be compared across States, over time, and against established standards. Second, it creates a data base to give program managers a foundation for improving the accuracy of certification.

Performance measurement and management information are closely linked. For instance, the information produced by QC, if properly classified, presented, and fed back to program managers, can be a vital factor in improving certification accuracy. The more closely that performance measurement and management information functions are integrated, the more effective such feedback can be. However, the sample size and data requirements of these two functions differ to some extent, making total integration problematic. That is, a process ideally suited for measuring performance may sacrifice some management information, and vice versa.

The relative emphasis on the performance measurement and management information functions of QC can also shift over time as changes occur in the size of the program, the problems it experiences, and perceived policy priorities. The current FSP quality control system emphasizes performance measurement, largely because the three factors cited earlier--program size, program growth, and the allocation of fiscal responsibility for the program--have accentuated the importance of a systematic basis for program accountability.

QC reviews provide a basis for a system of financial liabilities and incentives that are designed to hold States accountable for certification accuracy. States that exhibit overpayment error rates beyond a specified threshold are liable for a portion of the cost of overpayments. Assessing

---

5/Negative action reviews focus on the correctness of the denial process, but do not establish the household's correct eligibility status or benefit amount.

liabilities provides a mechanism whereby both State and federal governments share in the cost of certification errors. Because food stamp benefits are fully funded by federal tax dollars, the federal government would bear the full cost of all erroneous payments in the absence of quality control liabilities. Quality control liabilities thus limit the fiscal burden imposed on the federal government by erroneous benefits, by redistributing some of the risk of erroneous certification decisions to State and local agencies. This allocation is reasonable since State and local agencies have the ultimate authority in individual eligibility and benefit determinations. Again, in 1985 alone, the cost of erroneous overpayments was nearly $900 million. Under the current system of liabilities, States were held accountable for less than 25 percent of this loss.

States that exhibit low rates of overpayments, underpayments, and improper denials or terminations are eligible for additional federal funding. This creates additional incentives to reduce error rates, even for States that are not subject to liabilities. In 1985, three States qualified for a total of $1 million in incentive payments.

B. DEVELOPING THE STUDY AGENDA

The QC study undertaken by the Department of Agriculture responds to the Congressional mandate in the context of the present QC system, the emphasis it places on performance measurement and liability, and the issues of most common concern that have emerged in recent years from operating the system. Public debate in recent years has focused on (1) the technical soundness of the QC performance measurement process, and (2) policy decisions which have increased the pressures exerted by the QC system to reduce errors, the number of States that face liabilities based on their error rates, and the amount of those liabilities. In developing its study plan, USDA has thus focused on the Congressional question about the reasonableness of the data generated by the QC system as a basis for financial liabilities. However, the Department also investigated the extent to which the QC system provides management information to States and how this function might be improved.

To develop and refine the scope and subjects of its study, USDA began its work with a broad search for areas of concern. The Department consulted with experts on quality control policy and operations both inside and outside the federal government. Department staff met with representatives of the Department of Health and Human Services, the Office of

Management and Budget, the National Council of State Human Service Administrators' Management and Quality Control Committee, the Technical Advisory Group of the Food Stamp Quality Control Directors, and the Association of Public Welfare Administrators. The Department also re-examined the legislative and regulatory history of QC and reviewed published critiques or commentaries on the QC system.

The initial search identified a large number of potential areas for study. Many, however, were only remotely relevant to assessing the current QC system requested by Congress. In keeping with the Congressional request, USDA narrowed the scope of the study and kept it within manageable dimensions, by focusing on the key issues most clearly linked to the Congressional mandate. Two basic premises guided the selection of issues and the definition of the study's scope.

The first premise is that the primary function of QC is to help ensure program accountability and, more specifically, to provide a measure of the accuracy of certification. The direction of the Congressional mandate is explicit--to study QC as the basis for withholding funds from States that exhibit excessive error rates. While there may be room for debate about the definition of error, the procedures for measuring errors, and the manner in which liabilities are assessed, it is clear that Congress also views QC liabilities as a vehicle for limiting the fiscal burden imposed on the federal government by erroneous payments.

Therefore, this study does not deal with broader measures of program performance (such as administrative efficiency and the timeliness of application processing and other administrative actions). Each of these measures is important in its own right, but none is intrinsic to the objectives and operations of QC as defined by Congress. As noted earlier, each is reviewed under the broader Performance Reporting System. Moreover, both the General Accounting Office--the nonpartisan investigative arm of Congress--and USDA's Office of the Inspector General conduct routine and special audits. Thus, QC should be viewed as only one component of a more extensive set of procedures whose purpose is to make the Food Stamp Program accountable to its legislative objectives.

The second premise used to define the research agenda is that the correct measure of program performance for accountability purposes is the accuracy of the outcome of certification, rather than strict adherence to procedural requirements. States have the flexibility to tailor administrative procedures to their own needs and preferences within the policy framework set by the federal government. Given the

wide variation in State and local settings, there is every reason to believe that very different procedures can lead to the same desirable outcome--in this case, certification accuracy. In fact, different administrative approaches may be not only desirable but also necessary. What works well in a densely populated, large urban area may not work in a small rural setting. Consequently, the variety of administrative practices is extensive.

This variety is an appropriate response to the division of administrative responsibilities among federal, State, and local governments. However, it is of only secondary interest to determine whether or not a State or local office correctly follows its administrative procedures. The question that QC seeks to answer is whether or not benefits are paid correctly. The adherence of State and local offices to any given set of administrative procedures may or may not generate accurate payments. Thus, it is possible to achieve process-oriented goals without achieving important administrative outcomes. A measurement system that focuses on an administrative process rather than outcomes can create a distorted view of program performance.

Within this context of the accuracy of the outcome of certification decisions, the study focused on four characteristics of a sound QC system:

1. The QC system should be balanced. Achieving the goals of quality control requires that all relevant aspects of certification error be measured. Therefore, the study has examined the types of errors that are measured and the manner in which alternative definitions of error or the scope of the error-rate measure can create different performance incentives.

2. The QC system should be accurate and reliable. The study has examined the appropriateness and technical correctness of the methods used to select sample cases, conduct reviews, and calculate error rates. The statistical procedures used in QC have been examined in great detail to assess their ability to generate unbiased and reasonably precise error rates.

3. The QC system should be fair. An important focus of the study has been to determine whether the present system affects States in similar circumstances comparably.

4. The QC system should provide clear and understandable incentives. It is important to minimize the mysterious-

ness of the QC process and its results. Although some aspects of quality control are inevitably complex, the extent to which any alternatives reduce or improve the clarity of the process to program managers is an important concern.

The specific issues examined in the report that follows fall into two general categories: largely technical issues for which empirical analysis can inform the decision process, and policy issues for which there are no strictly technical answers. The first set includes such issues as the statistical properties of the estimated error rates, and the effects that characteristics of the program environment, independent of State administrative efforts, have on error rates. USDA reviewed the body of existing research and sponsored important new research efforts to shed light on these issues.6/

The second class of issues is less easily illuminated by empirical analysis and cannot be resolved with strictly technical answers. These issues are examined in full recognition that the current QC system reflects policy choices, and that a re-evaluation of policy choices over time is appropriate. These issues include questions about the acceptable degree of sampling error, the extent to which State and federal governments should share in the cost of certification errors, and the manner in which error-rate results should be used to allocate the cost of errors between the States and the federal government.

The study addresses many questions for which there are no right or wrong answers. In the end, they require policy decisions. To help guide the policy debate, USDA has highlighted the importance of the issues and explored the advantages and disadvantages of the current system's design. On some issues, the weight of the evidence enables USDA to reach clear conclusions. In other areas, the appropriate approaches are less obvious, and more than one alternative may be acceptable. In these areas, USDA does not make any recommendations. Instead, the Department plans to continue its deliberations, including a thorough review of the parallel report prepared by the National Academy of Sciences under contract to USDA.

6/Individual reports on these research efforts will be issued in the near future. Their conclusions are summarized in this report.

C. ORGANIZATION OF THE REPORT

The Department of Agriculture's report to Congress consists of six chapters. Chapter II provides a detailed description of the current QC system, as well as a brief history of its development. Chapters III and IV respond to the Congressional concern about the quality of QC data and the reasonableness of the methods for using these data to establish financial liabilities. Chapters III and IV, in broad terms, address technical and policy issues, respectively. Chapter III deals with issues that pertain to the definition of error, procedures for measuring errors, and the capacity of current procedures to yield accurate measures of performance, without bias or serious imprecision. Chapter IV focuses on the application of error-rate results and the associated policy choices: how to set error-rate thresholds beyond which States are liable for the cost of errors, and how to compute the relative shares of the cost of errors that should be paid by States and the federal government. Chapter V responds to the Congressional concern about the current and potential uses of the QC system as a source of management information to help State agencies improve their programs.

Throughout the chapters that follow, however, the reader should keep in mind that, while the report analyzes individual aspects of the quality control process, the components of the system are closely interrelated. The system's balance, accuracy, fairness, and clarity are the result of the net effects of all technical and policy decisions. More precise technical solutions, for example, may severely complicate the overall process, make it less comprehensible, and in the end have little effect on incentives or performance. Modifying the scope of errors counted in the QC system could create a more balanced set of incentives, but it might also sharply alter the reported incidence of errors and require a reconsideration of how the costs of errors are shared. Chapter VI summarizes the conclusions reached in Chapters III through V on these various issues.

II. DESCRIPTION OF THE QUALITY CONTROL PROCESS IN THE FOOD STAMP PROGRAM

The legislative and executive branches of the federal government have broad responsibilities for defining, supporting, and monitoring the Food Stamp Program. The federal government defines Food Stamp Program policy in the form of legislation, regulations issued by the Secretary of Agriculture, and policy handbooks issued to State agencies. It provides most of the funding to support the program; it pays for the full cost of benefits issued to participating households and for approximately half of the administrative costs of running the program at the State and local levels. The Food Stamp Act of 1977, as amended, requires the federal government to take specific steps to ensure that these funds are spent correctly. The Secretary is enjoined to "establish standards for the efficient and effective administration of the food stamp program by the States" and to "institute an error rate reduction program under which, if a State agency's payment rate exceeds [5 percent] the Secretary shall . . . reduce the State Agency's federally funded share of administrative costs." As a result of this legislative requirement, the Department of Agriculture operates a Performance Reporting System (PRS) and, as part of PRS, a quality control system to measure the accuracy of the certification processes of States and to hold States accountable for administering the Food Stamp Program effectively and accurately.

This chapter describes the current quality control (QC) process and its historical evolution. The description provides important background for the discussion in Chapters III, IV, and V on issues associated with error definitions, error measurement, error-rate targets and financial liabilities, and management information obtained from the QC system. Section A first presents a brief history of the QC system, and Section B broadly outlines the current system as it has emerged. Sections C through F then describe the current QC system in detail.

A. A BRIEF HISTORY OF FOOD STAMP QUALITY CONTROL

The current food stamp quality control system reflects over 15 years of policy debate, technical scrutiny, methodological development, and legislative refinement. The major features of the current system, thus, resulted from considerable experimentation and experience. Many of the issues examined in detail in later chapters have emerged repeatedly during

past reviews of the QC system in its early stages. Current practice reflects the resolution of competing objectives and improvements in technical procedures achieved in these earlier stages. This 15-year development occurred in three major phases of regulatory and legislative actions that have affected quality control: the first regulatory implementation of QC in 1971, the Food Stamp Act of 1977, and the 1980 and 1982 amendments to the Food Stamp Act of 1977.

1971 Regulatory Implementation of Quality

The original QC regulations, implemented in November 1971, preceded legislation by six years, although they did reflect Congressional concerns. Quality control was viewed by program managers as a way to improve program audits that were already in use, to enhance the ability of agencies to monitor program operations, and to provide a basis for developing corrective action plans when necessary.

Even this earliest set of QC procedures called for reviewing a statistically reliable sample of households and comparing State error-rate performance with defined targets. Beginning with methods that were already then being used by the AFDC program, this first food stamp QC process used a combination of error measures. Error rates were compared with two targets: a 3 percent target for eligibility errors and a 5 percent target for "basis of issuance" errors--error amounts involving overissuance or underissuance to eligible households. Reviews were conducted on both active cases and negative actions (the latter entail decisions to deny an application or terminate a household's participation), drawn from the non-public assistance caseload. Sample sizes similar to those currently used were prescribed for 6-month review cycles (described later in Section C). Error rates based on the sample reviews were published starting in 1974, and served as a benchmark for joint efforts by the federal government and States to identify problems and methods for reducing errors.

Several changes and improvements were made in the QC process over the ensuing five years. Statistical consultants helped the Food and Nutrition Service improve its criteria for reviewing State plans, and the training of federal and State staff increased the procedural consistency of QC reviews. In 1975, quality control results were formally made part of the new comprehensive Food Stamp Program Performance Reporting System, and, in 1976, the QC process was broadened to include public assistance households.

The Food Stamp Act of 1977

The Food Stamp Act of 1977 established the first legislative requirement for a quality control system. The law established increases in administrative funding for States that exhibited low error rates and a new definition of error for determining eligibility for this higher match rate. The Act provided an administrative cost reimbursement of 60 percent, rather than the normal 50 percent rate, for States whose combined overpayment and underissuance error rate was below the 5 percent basis of issuance target. States which failed to meet this target were required to develop corrective action plans.

These new legal consequences of error-rate performance increased the importance of a sound technical design for the QC system. Thus, USDA suspended the existing QC system between August 1978 and September 1979 to devote full attention to implementing the 1977 Act. The new system, initiated in October 1979, contained a uniform sampling methodology and increased the size of samples that were re-reviewed by federal staff. At this time, FNS began work on a system to provide automated processing of QC data. FNS also removed from the error calculations any remaining procedural deficiencies which did not affect benefit amounts.

Food Stamp Amendments of 1980 and 1982

The persistently high rates of error in the Food Stamp Program led to concerns that an even stronger system of incentives was required to improve certification accuracy. In response, the 1980 amendments to the Food Stamp Act established the concept of assessing financial liabilities against States that exhibited high error rates and providing incentive funding for States with low overall error; performance was compared with a standard based on the average performance among States and improvement in error rates. States which failed to meet their targets faced liabilities equal to the difference between the cost of the errors (overissuances plus underissuances) and their targets. Liabilities for the first sanction period were waived by the Assistant Secretary of Agriculture after the States submitted special correction action plans.

Further changes were initiated by the 1982 amendments which established the process now in place. These amendments:

- Eliminated underissuances from the payment-error rate

- Set error-rate targets at 9, 7, and 5 percent for fiscal years 1983, 1984, and 1985[7]/
- Changed the period for performance review from 6 months to a full year
- Made administrative costs rather than benefit issuances the basis for assessing financial liabilities
- Revised the incentive funding formula

B. OVERVIEW OF QUALITY CONTROL SYSTEM

The current QC process that has evolved from these successive refinements is conducted in two major stages: the State review and the federal re-review. In the State review process, samples of food stamp households are selected, and State QC staff conduct intensive reviews of each case to determine whether the eligibility and benefit decisions recorded in the case file were based on an accurate assessment of household circumstances and the correct application of food stamp policy. The results of State reviews for each case are recorded in QC review files and transmitted to the Food and Nutrition Service (FNS) of the Department of Agriculture. Based on the State review results, a reported error rate is computed.

The federal re-review establishes an official error rate, which forms the basis for assessing liabilities and offering incentives for State performance. Federal QC staff in each FNS Regional Office select a subsample of the cases reviewed by State QC staff, and conduct their own review to determine whether household circumstances were correctly evaluated and eligibility policy correctly applied. The results of the federal re-review are recorded for each case in this subsample; these results may differ from the results of the State review for particular cases. A statistical procedure called the "regression estimator" is then used to estimate for each State the error rate which would have been found had both the federal and State reviews been performed for the entire State

---

7/For FY 1983 and 1984 States with historically high error rates could avoid liabilities by improving their error rates significantly even if they did not reach the 9 and 7 percent targets.

caseload.8/ This regressed error rate is then adjusted by the degree to which the State failed to complete reviews on all its sampled cases; this last step yields the official error rate.

The official error rate is then compared with a Congressionally mandated target or threshold. States are financially liable for error rates that exceed the threshold. States whose error rates are below the threshold can receive incentives in the form of enhanced federal reimbursements under certain circumstances.

The remainder of this chapter describes this process in greater detail. Section C describes the process whereby States select their review samples, and Section D describes the process whereby they conduct reviews and report their results. Section E describes the federal re-review, and Section F explains the procedure for determining the official error rate based on federal and State review results. Section G presents the details of the process whereby financial liabilities and incentives are determined.

C. SAMPLING FOR STATE REVIEWS

The QC process measures error rates by reviewing a sample of active food stamp households and a sample of negative actions in which household benefits are denied at application or subsequently terminated. The accuracy of error rates estimated from these samples depends on rigorous adherence to valid sampling methods. FNS therefore establishes certain guidelines and requirements pertaining to the sampling process, which cover (1) the submission of a State plan, (2) the size of the samples to be drawn, (3) the sampling methods to be used, and (4) the timing of the sampling and review process.

Submission of a State Sampling Plan

Each State must submit a QC sampling plan for approval by FNS. In this plan, the States must describe the "sample frame," or the caseload pool from which they will draw their samples, and the method whereby they will select cases. If

---

8/The regressed error rate is not a correction of State review results. It is, instead, a separate measure derived from the federal subsample. The correlation between federal and State findings is used to enhance the reliability of the federal estimate.

States choose to stratify their samples, they must describe the procedures whereby they will divide the caseload into strata and the method whereby they will weight the error-rate results from each stratum to arrive at a final error rate for the caseload as a whole.9/ The State must define its target sample size according to federal requirements and various "schedule" options made available to the State (explained more fully below). Essentially, the State plan documents the State's adherence to specific federal requirements, as well as the methods chosen by the State in areas where federal guidelines allow the States some latitude.

Sample Size

Federal guidelines on the size of QC samples reflect a balance between the desire for statistically precise error-rate estimates and concerns about the cost and burden on State agencies of conducting the State-level review. Thus, federal rules prescribe several "schedules" of minimum required sample sizes that distinguish the sample sizes required for States with small and large caseloads, and the sample sizes required for reviews of active households and reviews of negative actions (denials and terminations).10/

Sample Size Requirements for Active Households. Federal rules base minimum sample sizes for active household reviews on the size of a State's average monthly food stamp caseload, and also allow the State to choose between a "high schedule" and an alternative "low schedule." The basic requirement is expressed in the high schedule, which links the size of the

---

9/The sample is stratified by dividing up the overall caseload into segments (strata) based on specified characteristics (e.g., public assistance households and non-public assistance households), and drawing sample cases separately from each stratum.

10/States may choose to review samples larger than the required minimum, in order to increase the precision of State error-rate estimates or allow estimates to be developed for sub-state administrative units. Federal administrative reimbursement is provided for such extra reviews at the same rate as for the basic required sample.

annual QC sample of active households to the State's average monthly caseload as follows:11/

BASIC SAMPLE SIZE REQUIREMENT FOR ACTIVE HOUSEHOLDS

Average Monthly Caseload (N)	Annual Sample Size (n)
60,000 or more	2,400
10,000-59,999	300 + .042(N - 10,000)
Under 10,000	300

However, States may use an alternative "low schedule" to reduce their cost of performing QC reviews. States which use this lower schedule must agree in their State plan not to challenge the final QC error rate on the grounds that the sample does not provide sufficient precision to ensure that the true error rate is indeed above the liability threshold. The alternative schedule available to the States also links sample size to the average monthly caseload, as follows:

11/Sample size requirements are set according to each State's forecast of its average monthly caseload in the coming year. QC policy recognizes, however, that actual caseloads may differ from such estimates, and allows considerable leeway without adjustment of the sample size requirement. If a State's caseload exceeds its forecast by less than 20 percent, no adjustment in sample size is required. When the differential is higher than 20 percent, some States reduce their sampling rate in the latter part of the year to keep the total sample size at the level they had planned for, thus avoiding the increased costs of reviews.

ALTERNATIVE SAMPLE SIZE REQUIREMENT FOR ACTIVE HOUSEHOLDS

Average Monthly Caseload (N)	Annual Sample Size (n)
60,000 or more	1,200
10,000-59,999	300 + .018(N - 10,000)
Under 10,000	300

The distinction between the high and low sample-size schedules stems from a legislative change made in 1982. At that time, Congress directed USDA to base liabilities on the results of an annual review period rather than the semiannual periods used up to that time. Larger States had been selecting and reviewing cases at an annual rate of 2,400 (based on the maximum semiannual sample of 1,200). Thus, the switch from semiannual to annual review periods enabled States to double their effective sample size at no additional review cost. However, in recognition that the larger sample size would generate only a modest increase in the precision of the estimated error rates, states were given the option of reducing the total number of cases they reviewed each year so as to reduce their review costs. In 1985, 31 states elected to use the lower schedule. Table II.1 shows the average monthly caseloads and the required sample sizes of all States in 1985.

Sample Size Requirements for Negative Actions. The required number of reviews for negative actions (benefit denials and terminations) has been lower than that for active cases since the start of the FSP quality control system in 1971. As described in more detail in later sections, the negative action review focuses on whether or not the decision to deny or terminate benefits was procedurally correct; it does not determine whether eligible households were wrongly denied benefits. (Answering that question would necessitate establishing whether, had procedures had been followed, the household would have been found eligible).12/ Since the aim

12/States are required to restore erroneously denied benefits. This restoration, however, lies outside the QC system.

Table of Contents

TABLE II.1

ACTIVE SAMPLE SIZE REQUIREMENTS: FY 1985

State	Average Monthly Caseload	Required Active Sample Size
Alabama	207,036	1,200[a]
Alaska	7,690	300
Arizona	62,359	2,400
Arkansas	92,000	1,200[a]
California	493,320	2,400
Colorado	64,492	1,200[a]
Connecticut	53,052	1,075[a]
Delaware	13,700	367[a]
Dist. of Columbia	29,580	652[a]
Florida	237,654	2,400
Georgia	202,755	1,200[a]
Guam	4,686	300
Hawaii	35,491	759[a]
Idaho	19,239	522
Illinois	349,273	2,400
Indiana	149,898	1,200[a]
Iowa	75,827	1,200[a]
Kansas	44,810	927[a]
Kentucky	184,071	1,200[a]
Louisiana	202,381	1,200[a]
Maine	46,850	963[a]
Maryland	114,038	1,200[a]
Massachusetts	139,010	1,200[a]
Michigan[b]	414,216	1,800[a]
Minnesota	90,000	1,200[a]
Mississippi	165,116	1,200[a]
Missouri	137,000	2,400
Montana	21,430	574
Nebraska	33,942	875
Nevada	13,359	381
New Hampshire	12,388	357
New Jersey	157,458	2,400
New Mexico	50,327	1,026[a]
New York	752,209	1,200[a]
North Carolina	174,137	1,200[a]
North Dakota	12,100	317[a]
Ohio	427,532	2,400
Oklahoma	100,600	1,200[a]
Oregon	93,886	2,400
Pennsylvania	414,858	1,200[a]
Rhode Island	29,033	757
South Carolina	131,694	1,200[a]
South Dakota	15,315	428
Tennessee	185,413	1,200[a]
Texas	392,974	1,200[a]
Utah	25,503	570[a]
Vermont	17,021	469
Virgin Islands	7,969	300
Virginia	138,323	1,200[a]
Washington	105,454	2,400
West Virginia	94,161	1,200[a]
Wisconsin	112,659	2,400
Wyoming	9,297	300
U.S. TOTAL	7,164,585	63,219

[a]State elected to use a reduced sample size in accordance with section 275.11(a)(2)viii of the Food Stamp regulations.

[b]FNS granted Michigan a special adjustment to reduce its sample size to 1,800.

of these reviews is to identify potential problems rather than to measure incorrect issuance or establish a basis for liabilities, the lower level of precision inherent in the smaller sample size is accepted. Thus, QC resources are focused to a greater extent on samples of active households. A single schedule of sample-size requirements applies to all States for negative action reviews, linking annual sample sizes to the average monthly number of actions:

SAMPLE SIZE REQUIREMENTS FOR NEGATIVE ACTIONS

Average Monthly Caseload (N)	Annual Sample Size (n)
5,000 or more	800
500-4,999	150 + .144(N - 500)
Under 500	150

Table II.2 shows the average monthly number of negative actions and the required sample sizes for negative action reviews undertaken by the States in 1985.

Sampling Methods

Federal QC policy governs the two basic steps in sampling: establishing the sampling frame and selecting review cases. The first step in sampling is to establish the sample frame, or, again, the pool of households from which review cases will be drawn. Although sample-size requirements are stated in annual terms, sample frames must be identified and a portion of the sample must be drawn each month, since QC policy requires that reviews be performed promptly after the "sample month," the month in which the benefit issuance or action in question is to be reviewed. Two separate sample frames must be constructed--one for active households, and one for negative actions.

Both sample frames are defined by a process of identification and exclusion. The active sample frame consists primarily of households that received food stamps during a particular sample month, and the negative action sample frame consists of all households whose application for food stamps was denied or

Table of Contents

TABLE II.2

NEGATIVE ACTION SAMPLE SIZE REQUIREMENTS: FY 1985

State	Average Monthly Number of Negative Actions	Required Negative Action Sample Size
Alabama	15,312	800
Alaska	1,914	354
Arizona	5,716	800
Arkansas	7,100	800
California	53,388	800
Colorado	6,200	800
Connecticut	1,202	251
Delaware	856	201
Dist. of Columbia	690	177
Florida	9,945	800
Georgia	16,024	800
Guam	238	150
Hawaii	1,641	314
Idaho	3,670	558
Illinois	17,921	800
Indiana	5,621	800
Iowa	12,146	800
Kansas	4,468	699
Kentucky	9,050	800
Louisiana	8,637	800
Maine	3,688	609
Maryland	2,917	498
Massachusetts	3,247	546
Michigan	29,839	800
Minnesota	7,400	800
Mississippi	6,250	800
Missouri	10,000	800
Montana	1,400	280
Nebraska	1,818	340
Nevada	3,117	527
New Hampshire	1,043	228
New Jersey	8,128	800
New Mexico	2,337	415
New York	27,086	800
North Carolina	11,986	800
North Dakota	870	203
Ohio	9,538	800
Oklahoma	10,000	800
Oregon	3,630	601
Pennsylvania	3,024	513
Rhode Island	1,600	308
South Carolina	5,006	726
South Dakota	1,341	271
Tennessee	8,811	800
Texas	30,097	800
Utah	3,922	610
Vermont	1,357	273
Virgin Islands	96	150
Virginia	8,117	800
Washington	8,375	800
West Virginia	8,675	800
Wisconsin	10,379	800
Wyoming	1,317	240
U.S. TOTAL	418,150	31,642

whose certification was terminated during the sample month.13/ Federal QC policy gives States latitude to define procedures for stratifying and selecting their samples. States may choose to stratify their samples--that is, to divide the sampling frame into groups based on defined characteristics, and to sample separately from each stratum--to ensure that certain types of households are represented, to ensure that a minimum sample size is established for certain types of cases, or to facilitate integrating QC reviews for households which receive both food stamps and AFDC.14/ States may use either systematic sampling--choosing cases at fixed intervals--or random sampling to select sample cases from the sampling frame. Table II.3 summarizes the number of States that use stratified samples and integrated reviews.

Timing of State QC Reviews

Completing the State QC review process promptly is desirable for four reasons. By initiating reviews promptly, QC staff may be better able to locate the selected households for interviews. Second, prompt reviews increase the likelihood that reviewed households, when interviewed, will be able to provide accurate information about their circumstances in the sample month. Third, when reviews are undertaken promptly, collecting corroborating information from other agencies or institutions (e.g., banks or employers) is likely to be less problematic. Finally, the more rapidly QC staff initiate and complete reviews, the sooner information on program performance will be available.

For these reasons, QC policy places deadlines on the State review process. States are required to conduct reviews and report results promptly. The review results for all cases selected in a given sample month must be reported to FNS within 95 days after the end of the sample month. The monthly status report on all cases is submitted within 105 days after the end of the sample month.

13/However, certain households are excluded from review if all members of the household had died or moved out of the State since the sample month, or under certain other conditions.

14/Integrated reviews enable States to obtain both an AFDC and an FSP QC review from a single household which participates in both programs, thus reducing review costs. A sample stratum is defined which consists of AFDC/food stamp households, and the sample cases that are drawn from this stratum are used for the QC purposes of both programs.

Table of Contents

TABLE II.3

STATES USING INTEGRATED QC SAMPLES AND DISPROPORTIONATE STRATIFICATION: FY 1985

		Disproportionate Stratification[a]		
		Yes	No	Total
Integrated QC Samples	Yes	11	15	26
	No	4	23	27
	Total	15	38	53

[a]Disproportionate stratified samples are those in which different sampling rates are used in different strata.

D. CONDUCTING THE STATE QC REVIEW

After the review cases are selected each month, the following 4-step State review process begins: (1) a case file review, (2) a field investigation, (3) a determination of the case's error status, and (4) the reporting of QC data and results to FNS. Federal regulations and handbooks guide the process and help ensure that States follow uniform procedures.

For the most part, these functions are carried out by a designated QC unit in the State agency. QC reviewers may be housed in a central State office or located in field offices throughout the State to minimize travel for field investigations. However, QC reviews may not be performed by members of local Food Stamp Program project area staff, since maintaining the objectivity of QC requires that reviewers not have any previous knowledge of the cases they examine.

File Review

The first step in the QC review is an examination of the case file--the desk review. For both active and negative action households, the reviewer begins by examining the information in the case file and the specific facts that support the household's eligibility (or ineligibility) and the amount of benefits issued. The desk review (and the later field investigation) entails examining each "program element," or factor, used to determine eligibility and the allotment amount. These program elements are classified under four broad categories: (1) nonfinancial criteria (e.g., the composition and size of the household, and the disability, citizenship, residency, and school attendance of household members); (2) resources (e.g., bank accounts, vehicles, and personal property and real estate); (3) income (e.g., earnings and unearned income); and (4) deductions from income (e.g., excess shelter costs, child and dependent care, and standard deductions). Information about each program element is recorded on a QC worksheet, including pertinent facts, the sources of verification included in the file and their apparent validity, and any gaps or deficiencies in the contents of the case file.

If, using recipient-reported information, the reviewer can determine that the household was ineligible in the sample month, the data collection portion of the review is considered to be complete, without a field investigation. If, however, no conclusive evidence of ineligibility exists in the file (or some evidence comes from sources other than the household), the reviewer must continue with the field portion of the investigation.

Field Investigation

Field investigation entails administering personal interviews to the head of food stamp household or that person's spouse (or the household's authorized representative), as well as making collateral contacts (in person or by telephone) with banks, employers, landlords, and other sources that can potentially verify (or contradict) information found in the case file.15/ For active household reviews, the field investigation must include a personal interview (in the home if possible) with the household, as well as other collateral contacts as necessary. For negative action cases, interviews are not required, and household as well as collateral contacts may be made by telephone.

In either case, the further information collected in the field investigation is also recorded on the QC worksheet for each program element.

Determining Error Status

Using information collected in the desk review and the field investigation (if required), the reviewer can determine whether discrepancies or variances exist between information in the case record and the results the review finds for the same sample month. The reviewer completes a "computation worksheet" which shows, side-by-side, case information and the calculations of eligibility tests and benefit amounts as originally recorded and as based on the QC review data. This process identifies all differences in financial circumstances and in the amount of benefits.16/

Based on the identified variances, the QC reviewer then determines whether a case contains a payment error as defined by QC policy, and if so the amount of the error. Not all variances are treated as QC payment errors. For example, discrepancies of $5 or less in benefit amounts are excluded

---

15/Under the privacy laws of some States, collateral contacts may be made by the QC reviewer only if the head of the food stamp household signs a written release to authorize such contacts as part of the review process. However, refusal to cooperate with the review by not providing such authorization is grounds for the termination of benefits.

16/The amount of benefits is computed in the review process only for active households; QC procedures do not require that reviewers determine the amount of benefits that would have been issued to households whose applications were denied incorrectly.

from the payment-error calculation. Variances are not counted if they occur because verification was postponed for expedited service cases or if the sample month was still within the time period granted to the food stamp agency to change benefits on the basis of information which gave rise to the discrepancy. Variances that do not affect the benefit amount are also excluded from the formal QC error count. The reviewer summarizes the review by coding the formal QC error status of each case as "eligible with correct benefit," "overissuance," "underissuance," or "totally ineligible." For active households, the reviewer must also record the amount of the error if there was an over- or underissuance.

Reporting QC Results

The results of the QC review are reported to FNS on standard forms. For active households, the Integrated Review Schedule (which is also used to report AFDC and Medicaid QC results) is used.17/ On this form, the reviewer records case identification, summary data on household income and deductions, the identity and characteristics of each household member, and detailed information on the type and amount of income for each individual. The reviewer codes a description of the QC error findings, including the program element and sub-element ("nature code") to which the error pertains, the dollar amount of the error, and whether the error was due to a mistake by the agency or a misreport by the client. Data from completed review schedules are entered into computer files at the State agency and transmitted to FNS's Washington Computer Center.18/

For two reasons, not all sample cases end up as completed reviews. Some selected cases, upon examination, are discovered to be outside the scope of the target population, and are coded "not subject to review." In other instances, such as when a household refuses to cooperate, QC staff may be forced to code a case as "not complete."

17/A special negative action review schedule is also used.

18/In most States, an internal agency review of the QC reviewers' findings is conducted before results are forwarded to FNS, so as to check the accuracy of reviewers' work, and to ensure that errors are neither overlooked nor mistakenly identified.

E. THE FEDERAL RE-REVIEW

A second round of case reviews is undertaken by federal QC staff in FNS regional offices. This review monitors the accuracy of the State QC review process and its application of certification and QC policy. The results of the federal re-review, when combined with the State QC results, determine the official error rate. The federal re-review entails sampling from State review files, reviewing cases, and resolving disputes over differences between federal and State findings on individual cases.

Sampling for the Federal Re-Review

Federal re-reviews are made for a subsample of the active household reviews submitted by each State's QC unit and, under specified circumstances, for a subsample of negative action reviews. For active household reviews, the annual re-review sample is based on the size of the State's review sample, as follows:

FEDERAL SAMPLE SIZE REQUIREMENT FOR ACTIVE HOUSEHOLDS

Average Monthly Caseload (N)	Annual Sample Size (n)
1,200 or more	400
300-1,999	150 + .277(n - 300)
Under 300	150

Negative action cases must be re-reviewed only if overpayment, underpayment, and State-reported negative action error rates together are low enough to qualify the State for incentive funding (see Section G).19/ Some regions, however, routinely conduct a small number of re-reviews of negative action

---

19/The federal re-review of negative action cases is a desk review. Like the State negative action review, it examines the procedural correctness of the negative action, but does not attempt to determine the actual eligibility status of the household. Consequently, negative action errors do not necessarily imply that the household was entitled to benefits.

samples to monitor negative actions. The annual re-review sample for negative action cases is determined as follows:

FEDERAL SAMPLE SIZE REQUIREMENT FOR NEGATIVE ACTIONS

Average Monthly Caseload (N)	Annual Sample Size (n)
800 or more	160
150-799	75 + .130(n - 150)
Under 150	150

For both active cases and negative actions, federal re-review staff must review all cases in their subsamples which are drawn from the State review sample, and must also review all State sample cases coded as "not subject to review" or "not complete." The re-review thus examines not only the correctness of State review results, but also the correctness of case classifications for cases not directly counted in the State's reported error rate.

Performing the Re-Review

The Federal re-review focuses on answering three questions about each State review case:

1. Did the State reviewer apply certification policy correctly?

2. Did the State reviewer apply QC review procedures properly?

3. Were the recorded results and findings of the State review accurate?

The federal re-review begins with a desk review of State-reported findings and is extended, as necessary, to resolve issues. If the desk review indicates mistakes or an inadequate investigation in the State review, the next step is to verify questionable information by making telephone calls to the household and collateral contacts as necessary. Field trips to interview the household are made if necessary. After the re-review, each completed case is classified according to

whether the federal reviewer agrees with the State's finding, agrees but notes procedural deficiencies, or disagrees with the State's finding. The federal re-review arrives at a federal finding that the household was eligible with the correct benefit amount, that the household was totally ineligible, or that the household was eligible with a specified amount of overissuance or underissuance.

Regional FNS offices must review each case within 60 days after they receive the State review file (unless additional information is needed and the reasons for the delay documented). Once a federal case re-review is complete, the case file and federal results must be returned to the State within 7 days.

Resolving Differences

When the federal re-review of a case generates a different error status than found by the State review, States have recourse to an appeal process to contest the federal finding. First, the State may request formal arbitration by the FNS regional office. The finding is arbitrated by a person in the regional office who had no involvement in the re-review process for that State, and must be performed within 30 days after the State's request for arbitration. In six of the seven federal regions, a State may first initiate a preliminary, or informal, discussion with the regional office even before this formal arbitration step, to present further information about the case or present its interpretation of certification or QC policy underlying the State result.

If the regional office's position is unchanged after formal arbitration at the regional level, the State can appeal to the national office of FNS for a final review. At whatever point differences are resolved, that decision is used to determine official QC error rates.

F. DETERMINING THE OFFICIAL QC ERROR RATE

The accountability function of the QC system is based on an official QC payment-error rate for each State, encompassing overissuances to eligible households and issuances to ineligible ones. This official rate is used to compare State performance with legislated targets and to establish the financial liabilities of States whose error rates are above the targets, as explained in Section G. The official error rate is determined by each regional office for the States in its jurisdiction, using standard statistical software provided by the FNS national office.

The official error rate is derived from a two-step process which draws on the results of both the federal re-review and the State reviews. In the first step, a procedure called the regression estimator is used to derive the regressed error rate. This procedure entails using the results obtained from the State review sample and the federal subsample to estimate the error rate that would be derived if the federal QC review, preceded by a State review, were applied to the entire caseload. The regression estimator does not adjust State-reported error rates. Rather, by using the double sample results, it increases the precision of the error rate estimate relative to what could be obtained from the federal sample alone. In the second step, the regression estimate is adjusted according to the percentage of cases in the original State sample for which reviews were not completed. The purpose of this adjustment is to maximize incentives for completing reviews at the State level.

The Regression Estimator

The federal re-review provides an external, objective error measure that helps ensure that QC policy is applied rigorously and consistently across all States. The federal re-review alone could be used as a basis for estimating the payment-error rate, but the sample sizes in the federal review limit its precision. Given the availability of the results from the State review sample, however, it is possible to improve the precision of estimated error rates by using information from both samples.

The regression estimator method was developed to make use of the results of both samples. This approach has generally had the effect of reducing the variance of the estimated payment-error rates by about 50 to 75 percent, as compared with the variance of estimates from the federal subsample alone. In effect, using the double sample results and the regression estimator generally yields results equivalent in precision to what would be attained by increasing the federal subsample to

where

$\bar{y}$ is the average overpayment per case in the federal subsample, as determined by the federal review (the average over all sample cases, including error and non-error cases);

$\bar{X}$ is the corresponding average overpayment in the State sample as determined by the State review;

$\bar{x}$ is the average overpayment as determined by the State review for cases in the federal subsample; and

$\bar{u}$ is the average issuance for all cases in the State sample.

In the above equation, b is the regression coefficient estimated from the federal subsample--an estimate of the relationship between the average federally determined overpayment per case and the average State-determined overpayment per case for cases included in the federal subsample.

It is important to understand that this is simply a procedure for reducing the sampling error of the estimate from the federal subsample. It makes use of the fact that the federal and State findings on individual cases are highly correlated. Consequently, if the overpayment errors based on State findings for the cases in the federal subsample are above those in the full State sample, then the federal findings based on that sample are also likely to be too high. The regression estimator adjusts for the difference in average State findings in the two samples. A similar adjustment is made if the State findings in the federal sample are below the State findings in the full State sample. The effective sample size of the federal subsample is increased substantially with the regression estimator since there is a high correlation of case by case findings from the State and the federal reviews.

The effect of the State findings on the size of the sampling error thus depends on two factors:

1. The difference between the average overpayments found by the State in the full sample and in the cases included in the subsample.

2. The correlation between the federal results and the State results for the cases in the subsample. If the correlation between State and federal findings is perfect,

the effective sample size for the regression estimator is the same as the full State sample.

Adjustment for Incomplete Reviews

Failure to complete QC reviews undercuts the validity of the QC results in two ways. First, it reduces the effective sample size, and thus reduces the precision of reported error-rate estimates. However, more important is the substantial likelihood that cases for which reviews are not completed are systematically different from completed cases; thus, their exclusion from the sample may bias the estimated error rate. On the assumption (supported by some evidence from special studies) that cases for which reviews are not completed are in fact more likely to contain errors than the rest of the sample, QC procedures adjust the regressed error rate based on the percentage of sample cases not completed.

This adjustment imputes to incomplete reviews a higher error rate than is found in the rest of the sample. Under current procedures, incomplete reviews are assigned an error rate which is higher than the regression estimate of the error rate by an amount equal to twice the standard deviation of the payment-error rate computed in the State-level review. The official error rate is the weighted sum of the regressed error rate and this higher assigned error rate for incomplete cases. The regression estimate of the error rate is weighted by the percentage of the State sample cases that were completed, and the imputed rate is weighted by the percentage for which reviews were not completed.

G. LIABILITIES AND INCENTIVES

Food stamp legislation establishes liabilities against States with high error rates, and some financial "bonus"--or incentive payment--for States with low error rates. Under current legislation, both liabilities and incentives are based on the amount of federal reimbursement for State administrative costs in the Food Stamp Program. The following sections describe how liabilities and incentives are determined, and their history to date.

Determining Liabilities

QC penalties are determined in two steps: (1) deciding whether a liability should be assessed, and (2) calculating the amount.

The decision to assess liabilities is based on a comparison of a State's official payment error rate and an error rate threshold established by Congress. Beginning with Fiscal Year

1985, the Congressionally mandated threshold is 5 percent.20/ In establishing liability levels, the law includes only the official overpayment error rate. Thus, the QC system emphasizes holding States responsible for minimizing overissuances, although the underissuance and negative action error rates are relevant when incentive payments are considered.

For States whose official payment-error rate exceeds the threshold, a schedule links the liability amount to the State's federal reimbursement for the administrative cost of the Food Stamp Program.21/ This schedule is as follows:

SCHEDULE OF QC LIABILITY AMOUNTS

Percentage Points of Error beyond 5% Threshold	Liability as Percent of Federal Reimbursement
1.00% or less	5%
1.01%-2.00%	10%
2.01%-3.00%	15%
3.01% or more	15% + 10% for each additional percentage point or fraction thereof

Therefore, a State whose error rate was 9.5 percent in 1985 would face a liability equal to 35 percent of its normal administrative cost reimbursement--5 percent for each of the first three percentage points of error in excess of the 5

20/The basic threshold was set at 9 percent for 1983 and 7 percent for 1984, but State-specific thresholds were computed in both years, as described earlier in footnote 2.

21/The precise definition of the base is that portion of administrative costs which are reimbursable at 50 percent, in the absence of liabilities.

percent threshold, and 10 percent for each of the next two points or fractional points.22/

When States are formally informed of their error rate and liabilities, they are also advised that, under the law, liabilities can be reduced if State agencies can demonstrate that there was "good cause" for the high error rate, including such circumstances as natural disasters, strikes by agency employees, and significant growth in the caseload, factors which could be pointed to as introducing unforeseeable elements beyond the agency's control. State agencies have 30 days after receiving the initial notification of their official error rate and liabilities to request a good cause waiver. FNS reviews these requests and can grant full or partial waivers if the criteria are met.

After arbitration, notification of official error rates, and procedures for reviewing good cause waivers are complete, and after FNS has billed the States for any liabilities, States may also request an administrative appeal. This request must be filed within 10 days after receiving a federal billing for the liability. (Billings occur no sooner than 30 days after initial notices of error rates are received or shortly after decisions on good cause waiver requests are made, whichever is later). States must submit additional information to support the appeal within 30 days following the receipt of the FNS QC file from the appeals board. If requested by the States, hearings are held before a USDA appeals board outside the Food and Nutrition Service; otherwise the decision is based on the written submissions. The appeals board has up to 60 days from the receipt of the State's submission to schedule and conduct a hearing, and 30 days from the hearing to make a final decision. If the State disagrees with the determination resulting from this appeal, it can file suit against USDA in court up to 30 days after the administrative review determination. This overall process protects States from unreasonable decisions in the overall review process and the determination of liabilities.

Determining Incentive Payments

Current legislation calls for increasing the rate of federal administrative cost reimbursement to 60 percent for States that exhibit low error rates. A State qualifies for increased

---

22/Liabilities are capped, however, at an amount equal to the percentage points of the payment-error rate beyond the threshold, multiplied by the issuance in the review month.

reimbursement if (1) the sum of its overpayment and underissuance error rates is below 5 percent, and (2) its negative case error rate is lower than the national weighted average negative error rate in the previous year.23/

History of Liabilities and Incentives

Tables II.4 and II.5 provide State error rates for overpayments and underpayments for Fiscal Years 1980 through 1985. Table II.6 summarizes the status of liabilities for Fiscal Years 1981 through 1984, Table II.7 provides the incentive reimbursements for low error rates during the same years. (Liabilities for Fiscal Year 1985 are reported in Chapter 3). The bases for liabilities in these years were established under three successive pieces of legislation. Both official error rates and error-rate targets have declined since the first liabilities were assessed in 1981. However, because error rates did not decrease as much as did the targets, assessments have been increasing. For Fiscal Years 1981 to 1985, the Food and Nutrition Service has made a total of 144 assessments on 50 States for about $339 million. As of April 15, 1987, four of the 144 liabilities have been paid (three by Connecticut and one by Wyoming); 34 assessments involving 20 States have been waived in full by FNS under laws applying to the period from October 1980 to September 1982; five assessments have been overturned by the administrative appeals board; and 101 assessments involving 49 States are pending.

Of the 101 assessments that are pending, 84 have either just been announced (Fiscal Year 1985) or are under good-cause review within FNS (Fiscal Year 1984); 11 are being appealed to the administrative appeals board (Fiscal Year 1983); and six are in the courts (from Fiscal Years 1981 to 1982).

Summary of the Overall QC Process Schedule

The full cycle of the QC process, from the first review month in an annual sample through the assessment of financial liabilities and the possible initiation of a State request for judicial review, is lengthy. However, an estimate of each State's error rate based on the State reviews will be available as soon as the State completes a year's work and no later than three and a half months after the end of the fiscal

---

23/Present policy on incentives took effect in 1983. For 1981 and 1982, the enhanced reimbursement rates ranged from 55 to 65 percent, depending both on how much a State's error rates fell below the threshold and on the size of its error-rate reduction from the previous year.

Table of Contents

TABLE II.4

QUALITY CONTROL PAYMENT ERROR RATES

	FY80	FY81	FY82	FY83	FY84	FY85
Connecticut	9.97	13.78	12.73	12.80	7.11	7.04
Maine	9.18	8.09	8.49	8.37	6.74	7.91
Massachusetts	10.48	11.31	13.38	13.58	9.86	9.71
New Hampshire	8.70	12.57	16.29	9.99	8.18	4.42
New York	15.78	13.63	11.42	9.98	10.14	7.11
Rhode Island	13.50	10.50	8.90	8.90	7.08	8.00
Vermont	10.40	9.23	10.26	16.71	9.53	8.06
Delaware	9.84	7.45	6.40	4.94	6.31	7.17
District of Columbia	14.86	13.12	11.10	10.08	8.80	9.81
Maryland	14.60	14.22	9.70	7.12	6.99	7.37
New Jersey	8.61	9.40	8.68	7.95	7.46	8.50
Pennsylvania	8.91	9.56	10.87	10.37	10.53	9.36
Puerto Rico	8.46	9.71	N/A	N/A	N/A	N/A
Virginia	7.69	7.50	8.20	6.46	8.04	6.67
Virgin Islands	12.45	10.44	11.40	14.77	12.12	9.73
West Virginia	7.40	9.09	9.03	5.52	6.94	5.07
Alabama	8.29	7.36	5.72	6.92	13.31	13.50
Florida	8.83	12.85	10.25	10.02	8.94	6.71
Georgia	9.48	9.80	8.34	7.53	9.89	12.91
Kentucky	7.14	7.75	7.15	6.90	8.98	6.00
Mississippi	10.36	10.10	9.11	8.36	9.18	7.98
North Carolina	9.74	11.37	10.51	7.71	7.11	6.49
South Carolina	10.48	9.03	9.57	8.11	10.61	12.10
Tennessee	10.41	11.31	10.04	6.99	6.08	6.39
Illinois	9.95	8.50	8.93	7.23	8.31	8.16
Indiana	7.43	8.08	7.41	8.77	8.95	10.90
Michigan	10.28	9.30	8.99	7.70	6.46	7.35
Minnesota	6.62	7.65	8.37	7.92	9.77	9.51
Ohio	8.45	7.74	8.56	6.90	6.65	7.43
Wisconsin	9.61	10.24	11.40	8.27	9.60	8.00
Arkansas	7.48	9.17	9.64	8.77	9.75	7.88
Louisiana	9.35	10.45	9.71	9.45	10.15	9.76
New Mexico	13.16	13.34	12.85	11.44	11.61	8.83
Oklahoma	6.97	9.31	8.02	8.79	7.19	10.58
Texas	7.65	9.28	9.69	7.58	9.60	10.38
Colorado	8.62	12.87	15.07	12.63	10.72	8.48
Iowa	10.26	9.11	9.25	8.81	8.48	8.41
Kansas	10.39	11.12	9.69	9.11	7.36	8.16
Missouri	8.00	8.52	7.40	6.74	5.81	5.23
Montana	9.13	13.48	7.56	5.52	8.83	7.44
Nebraska	12.16	11.02	10.67	7.16	8.75	9.04
North Dakota	6.68	5.16	6.89	4.98	6.27	3.53
South Dakota	9.19	8.27	10.64	7.84	3.59	3.15
Utah	10.86	7.89	9.79	13.33	11.37	7.26
Wyoming	10.42	12.36	8.72	9.88	9.07	6.78
Alaska	11.88	23.23	20.80	13.86	9.14	13.53
Arizona	10.87	12.25	11.98	9.79	9.58	9.38
California	7.51	7.11	8.61	6.98	7.70	7.08
Guam	6.40	7.97	5.31	7.57	3.39	5.33
Hawaii	4.49	6.97	5.96	4.28	3.70	4.35
Idaho	10.29	9.49	8.32	8.48	6.96	5.16
Nevada	4.08	3.39	1.48	2.17	2.54	2.48
Oregon	9.20	8.99	10.14	9.94	7.82	9.41
Washington	8.10	8.49	9.82	10.07	9.24	9.50
U.S. Total	9.51	9.90	9.54	8.33	8.61	8.30

Table of Contents

TABLE II.5

UNDERISSUANCE ERROR RATES

	FY80	FY81	FY82	FY83	FY84	FY85
Connecticut	1.70	3.16	3.14	2.79	1.90	1.78
Maine	2.05	2.61	1.89	2.34	1.57	1.39
Massachusetts	1.65	2.48	2.71	1.87	2.03	1.61
New Hampshire	1.90	2.66	1.82	1.93	1.94	1.88
New York	4.11	3.63	3.03	3.21	3.23	3.44
Rhode Island	2.93	2.17	2.48	2.56	2.01	1.47
Vermont	1.91	1.72	2.32	2.00	1.71	1.43
Delaware	2.22	2.76	2.07	1.87	2.35	1.76
District of Columbia	3.75	4.79	5.84	3.09	3.23	2.93
Maryland	2.45	2.47	1.63	2.19	1.33	1.43
New Jersey	1.86	2.11	2.33	2.44	2.15	2.06
Pennsylvania	2.25	2.46	2.02	2.02	2.09	2.36
Puerto Rico	1.74	2.07	N/A	N/A	N/A	.00
Virginia	1.85	2.02	2.39	2.14	2.30	2.46
Virgin Islands	2.60	3.82	2.03	4.79	2.24	1.87
West Virginia	1.70	2.57	1.99	1.85	1.53	1.52
Alabama	1.89	1.97	1.82	1.96	2.98	2.06
Florida	2.43	2.30	2.69	3.10	2.48	1.87
Georgia	2.54	2.71	2.37	2.35	3.42	4.13
Kentucky	1.57	2.00	1.98	1.88	2.03	1.67
Mississippi	2.61	1.90	3.39	3.01	1.90	2.28
North Carolina	2.86	4.64	2.46	3.29	3.51	2.34
South Carolina	2.15	2.33	2.10	2.90	3.68	3.48
Tennessee	2.28	2.48	2.30	1.94	2.04	1.41
Illinois	3.55	2.96	2.03	2.41	2.92	2.42
Indiana	1.72	.89	2.33	2.06	1.74	1.52
Michigan	2.99	2.85	2.75	2.07	1.54	2.13
Minnesota	2.16	1.83	1.96	1.72	2.10	2.45
Ohio	1.34	1.75	1.56	1.38	1.98	1.24
Wisconsin	3.22	3.47	4.32	3.38	3.20	2.61
Arkansas	1.56	2.51	2.81	1.95	2.23	1.76
Louisiana	2.30	2.44	2.80	2.48	1.71	2.08
New Mexico	2.35	2.10	2.72	3.03	2.23	2.11
Oklahoma	2.31	2.76	3.61	3.40	3.45	4.21
Texas	1.75	2.14	2.24	2.40	1.67	1.97
Colorado	1.43	2.58	2.53	2.33	2.03	2.71
Iowa	1.99	1.46	1.66	1.97	1.54	1.42
Kansas	2.46	2.52	1.51	1.88	2.31	1.99
Missouri	2.12	2.07	2.41	2.28	1.98	1.43
Montana	1.62	2.28	1.72	1.32	2.16	2.00
Nebraska	3.58	2.01	2.86	2.36	1.94	1.56
North Dakota	1.10	1.97	1.36	.73	.64	1.19
South Dakota	1.44	1.70	1.50	1.10	.94	.88
Utah	2.22	3.59	3.57	2.52	2.77	1.72
Wyoming	1.02	1.16	1.25	1.98	2.69	1.83
Alaska	2.64	1.90	2.73	2.46	2.98	2.59
Arizona	3.02	3.77	2.74	3.28	3.39	2.48
California	3.06	3.16	3.00	3.83	2.81	3.16
Guam	.56	1.94	1.69	1.42	1.16	1.11
Hawaii	1.89	2.33	1.86	1.24	1.08	1.25
Idaho	2.16	2.00	2.10	1.64	1.82	.99
Nevada	1.75	1.01	.90	1.05	.16	.54
Oregon	1.84	1.82	2.73	2.56	1.93	1.97
Washington	1.30	1.53	1.70	1.60	2.65	2.28
U.S. Total	2.35	2.50	2.44	2.45	2.34	2.25

Table of Contents

TABLE II.6

STATUS OF QC LIABILITIES ASSESSED
FY 1981-1984

Region/State	10/80-3/81 Amount	10/80-3/81 Status	4/81-9/81 Amount	4/81-9/81 Status	10/81-3/82 Amount	10/81-3/82 Status	4/82-9/82 Amount	4/82-9/82 Status	FY 1983 Amount	FY 1983 Status	FY 1984 Amount	FY 1984 Status
NORTHEAST												
Connecticut	$1,171,675	Waived	$1,338,545	Paid	$110,816	Paid	$60,539	Waived	$570,153	Paid		
Massachusetts	$469,709	Waived			$1,063,272	Waived	$1,368,893	In Court	$2,796,743	Appealed	$2,321,093	Notified
New Hampshire	$633,143	Waived	$285,165	Overturned	$318,222	Waived	$435,810	Waived			$70,125	Notified
New York											$10,063,964	Notified
Vermont									$705,015	Appealed	$200,169	Notified
MID-ATLANTIC												
District of Columbia					$225,343	Waived					$235,823	Notified
Maryland	$580,954	Waived	$619,367	Waived								
New Jersey											$1,088,471	Notified
Pennsylvania					$1,619,419	In Court			$2,316,399	Appealed	$7,819,005	Notified
Puerto Rico	$2,307,700	Waived										
Virginia											$1,304,695	Notified
Virgin Islands									$226,264	Appealed	$259,762	Notified
SOUTHEAST												
Alabama											$9,221,622	Notified
Florida	$1,647,478	Waived	$3,801,937	Overturned					$181,223	Appealed	$2,116,453	Notified
Georgia							$42,221	Waived			$3,697,445	Notified
Kentucky											$1,395,355	Notified
Mississippi											$1,731,884	Notified
North Carolina	$3,573,681	Waived	$2,501,744	Waived							$261,982	Notified
South Carolina											$3,159,387	Notified
Tennessee	$1,887,380	Waived	$209,003	Waived								
MIDWEST												
Illinois											$2,844,492	Notified
Indiana											$1,361,069	Notified
Minnesota											$1,461,779	Notified
Wisconsin					$728,433	Waived	$1,858,761	Waived			$1,391,622	Notified
SOUTHWEST												
Arkansas											$1,144,268	Notified
Louisiana									$965,340	Appealed	$5,283,439	Notified
New Mexico	$107,462	Waived			$623,045	In Court	$449,393	Waived	$563,423	Appealed	$2,197,196	Notified
Oklahoma											$231,271	Notified
Texas					$83,044	Waived					$8,212,334	Notified

Table of Contents

TABLE II.6 (cont'd)

Region/State	10/80-3/81 Amount	10/80-3/81 Status	4/81-9/81 Amount	4/81-9/81 Status	10/81-3/82 Amount	10/81-3/82 Status	4/82-9/82 Amount	4/82-9/82 Status	FY 1983 Amount	FY 1983 Status	FY 1984 Amount	FY 1984 Status
MOUNTAIN PLAINS												
Colorado	$230,030	Waived	$821,180	Overturned	$2,026,280	Overturn	$1,399,115	Waived	$1,059,295	Part Waived, Part Appealed	$1,381,910	Notified
Iowa											$690,194	Notified
Kansas	$342,690	Waived									$107,893	Notified
Montana	$329,294	Waived	$207,489	Waived							$101,927	Notified
Nebraska					$198,907	Waived					$301,193	Notified
Utah							$576,696	Overturned	$1,453,012	Part Waived, Part Appealed	$1,306,988	Notified
Wyoming			$47,274	Waived					$33,420	Part Waived, Part Paid	$94,377	Notified
WESTERN												
Alaska	$1,207,420	Waived	$2,148,102	In Court	$370,108	Waived	$404,420	Waived				
Arizona	$2,437,415	Waived	$236,206	In Court							$1,199,017	Notified
California											$4,263,749	Notified
Guam			$40,185	Waived								
Oregon					$668,221	In Court			$341,284	Part Waived, Part Appealed	$443,825	Notified
Washington									$705,919	Appealed	$1,509,980	Notified
TOTAL ASSESSED	$16,926,031		$12,256,197		$8,035,110		$6,595,848		$11,917,490		$80,475,758	
TOTAL STILL PENDING[a]	$0		$2,384,308		$2,910,685		$1,368,893		$10,490,641		$80,475,758	

[a]Amount pending excludes liabilities waived or overturned. For FY 1983, four states had liabilities partially waived: Colorado ($58,850), Utah ($726,506), Wyoming ($18,567), and Washington ($37,920).

Table of Contents

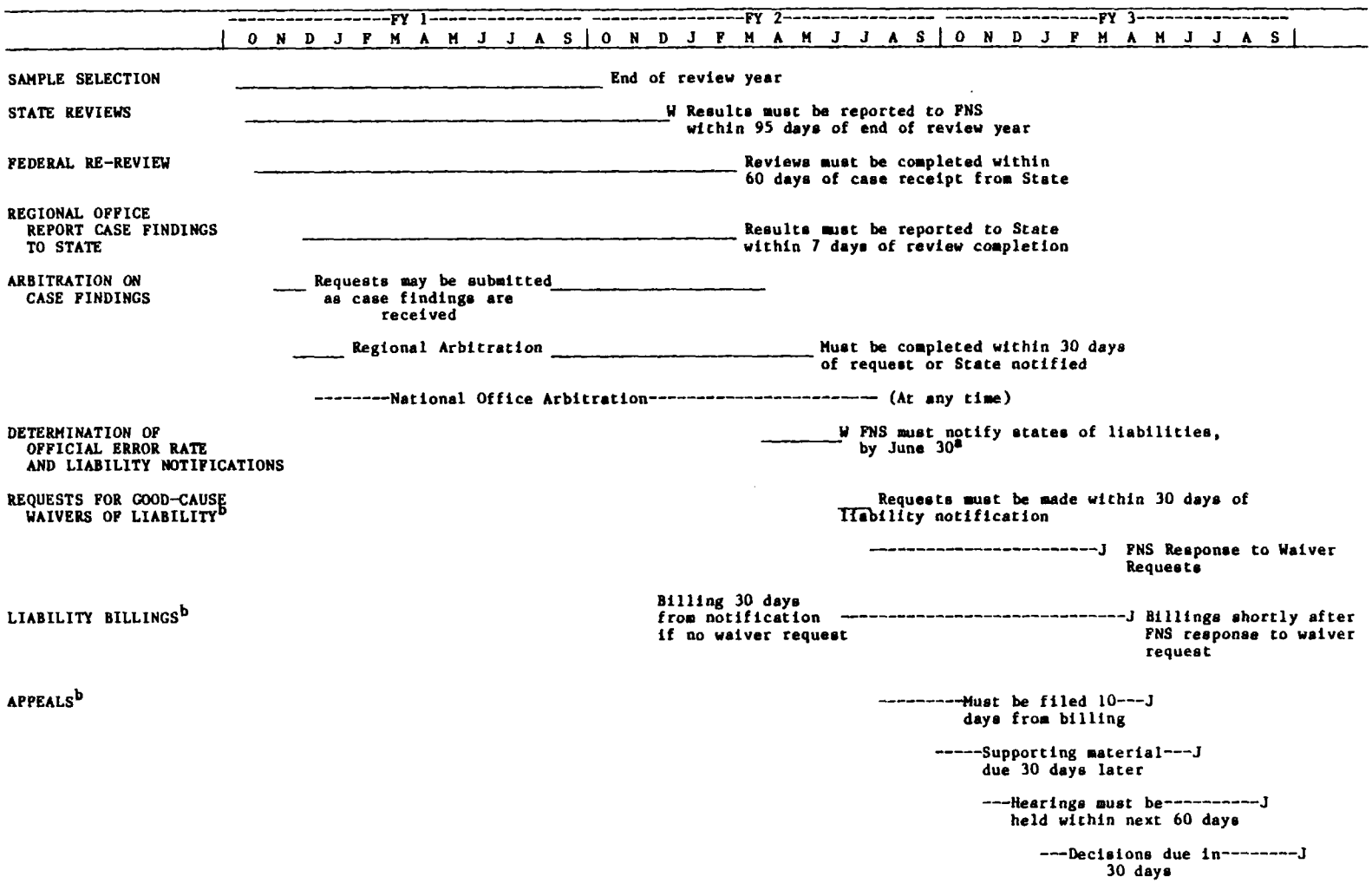
TABLE II.7
HISTORY OF INCENTIVE REIMBURSEMENTS

Region/State	10/80 - 3/81 Amount	10/80 - 3/81 Reimb. Rate	4/81 - 9/81 Amount	4/81 - 9/81 Reimb. Rate	10/81 - 3/82 Amount	10/81 - 3/82 Reimb. Rate	4/82 - 9/82 Amount	4/82 - 9/82 Reimb. Rate	FY 1983 Amount	FY 1983 Reimb. Rate	FY 1984 Amount	FY 1984 Reimb. Rate
NORTHEAST												
Rhode Island	$84,098	55%	$82,193	55%								
MIDDLE ATLANTIC												
Delaware					$60,796	55%						
Maryland							$634,554	55%				
Puerto Rico					$1,353,846	55%						
SOUTHEAST												
Alabama			$1,407,985	60%	$1,278,255	60%	$1,326,875	60%				
Georgia					$750,917	55%						
MIDDLE WEST												
Illinois	$698,267	55%										
Indiana			$836,878	60%								
Minnesota			$305,187	55%								
MOUNTAIN PLAINS												
Montana					$57,494	55%	$87,645	55%				
Nebraska	$76,395	55%										
North Dakota	$58,228	60%	$69,890	60%			$76,546	60%				
South Dakota			$133,535	60%							$345,725	60%
Wyoming					$23,795	55%	$27,187	55%				
WESTERN												
California			$2,143,062	55%								
Guam	$32,262	60%										
Hawaii							$204,477	60%				
Nevada	$173,994	60%	$309,340	65%	$316,885	65%	$343,035	65%	$383,252	60%	$398,876	60%
Oregon	$286,157	55%										
TOTAL	$1,406,436		$5,288,070		$3,841,988		$2,700,319		$383,252		$744,601	

year. Beginning with the results for Fiscal Year 1986, the federal official error rates must be announced within 9 months after the end of the review year. Figure II.1 summarizes the process described in this chapter and the deadlines that apply to each phase of the process.

Table of Contents

FIGURE II.1
TIMING OF QC CYCLE EVENTS



[a]Beginning with FY 1986 error rates and liabilities.

[b]Resolution of waiver requests may be a lengthy process, so the potential duration of FNS responses, billings, and appeals is shown as an open-ended period (---J).

III. TECHNICAL ISSUES: DATA QUALITY AND ALTERNATIVE ERROR DEFINITIONS

An important purpose of the quality control system is to provide a reliable, understandable measure of food stamp certification accuracy that can serve as a basis for federal efforts to hold State agencies accountable for their performance. This chapter of the report to Congress evaluates the food stamp QC system in terms of four technical and operational issues that affect the performance measurement capability of the system:

****

MAJOR TECHNICAL AND OPERATIONAL ISSUES

Issue A: Does the statistical design of the QC system ensure that results are unbiased and reasonably precise?

Issue B: Is the QC process operated accurately, reliably, and consistently?

Issue C: Does the definition of QC errors (i.e., what is counted as an error) form an appropriate basis for measuring performance?

Issue D: Should State QC error rates be adjusted in some manner to reflect differences in the operating environments of the States?

****

Chapter III thus focuses on largely technical questions about the ability of the system to maintain data quality, and its current operational approach to defining and measuring errors. These issues are classified as technical because they focus on operational methods and procedures. However, the distinction between purely technical and broader policy issues is not always clear-cut; for instance, some of the issues in this chapter involve both questions about the broad purpose and policy direction of the QC system and more technical issues about its procedural components. The policy questions that arise in connection with these issues are included in the discussion in this chapter. Chapter IV will deal with policy questions that pertain specifically to the methods used to establish an error-rate standard or threshold, and the approaches taken to determine the financial liabilities of States that fail to meet established standards.

This chapter examines each of the four major issues listed earlier. Each section begins with a list of more specific questions and a summary of the key findings. Each section then describes the analysis performed to address the specific questions, discusses the possible alternatives to current QC policy, and provides a final summary of the conclusions reached in each area.

A. THE STATISTICAL DESIGN OF THE QC SYSTEM

As described in Chapter II, official QC error rates are derived from an examination of a sample of food stamp households in each review period. The procedures for drawing quality control samples are methodological applications of standard and widely accepted sampling techniques. Sample surveys of this sort have considerable advantages over complete censuses of the population, in terms of their lower costs, greater speed, greater scope, and greater accuracy. However, the results of sample surveys are always subject to some uncertainty because only part of the population is examined.

Whether or not a sample will yield results which are sufficiently representative of the entire population depends primarily on whether the errors introduced by the sampling process are small enough for the task at hand. For the task of withholding federal funds from States that exhibit excessive levels of erroneous payments, the sampling and estimation process must be both unbiased and sufficiently precise.

Bias is a measure of the validity of an estimator.24/ An estimator is unbiased if the average value of the sample-based estimates, taken over all possible samples, is equal to the underlying population value. In other words, if estimates of State error rates are unbiased, their average over repeated samples will be very close to the true error rate.25/ A plot

---

24/The term estimator denotes the rule by which an estimate of some population characteristic is calculated from sample results. The term estimate refers to the value obtained for a particular sample.

25/In the context of food stamp quality control, the "true" error rate is the rate that would be found if State reviews and then federal re-reviews were conducted for the entire caseload.

of the distribution of estimates from repeated samples would be centered over the true value. An unbiased estimator is important to ensure that the QC process does not systematically overstate or understate the true error rate.

The sample estimates must also be reasonably precise. Estimates that vary substantially across repeated samples are less precise than estimates that exhibit little variation. The precision of an estimate is commonly stated in terms of its estimated sampling variance, or, equivalently, its estimated standard error. The smaller the standard error of an estimate, the greater its precision. Reasonably precise estimates are important to ensure that the risks of misallocating the financial costs of certification errors between the State and federal governments are acceptably low.26/

Statistical sampling theory provides useful guidelines on how the precision of estimates derived from random samples can be interpreted. If the sample size is known, and the estimated error rates follow what is called a "normal distribution"--that is, the familiar bell-shaped curve found in most introductory statistical texts--a single sample can provide a measure of the percentage of sample-based estimates that will fall within specified margins. This is commonly expressed as a "confidence interval"--a range around the sample estimate within which the true error rate falls within the computed interval for a specified proportion of all possible samples of the same size. That proportion is the "nominal" probability associated with the confidence interval. For example, for a particular State with a given sample size, an estimated error rate of 6.5 percent, and an estimated standard error of 1 percentage point, close to a 90 percent probability exists that the confidence interval from 4.9 to 8.1 percent includes the true error rate. In general, the width of a confidence

---

26/The question of how precise is "precise enough" is addressed in Chapter IV.

interval can be reduced--and the precision of the estimates increased--by increasing the sample size.27/

The key assumption in this discussion is that the estimated error rates do in fact follow the normal distribution (or bell-shaped curve). This assumption is a particular concern, since the underlying event of interest--a food stamp payment error--is a relatively infrequent occurrence. Over three-quarters of all food stamp cases are paid correctly. Of those with payment errors, some have relatively small errors, and some have quite large errors. Thus, the underlying distribution of errors for individual cases is highly skewed. However, the samples in many surveys are often large enough that estimates made from them are approximately normally distributed even if the underlying distribution is not. If this holds true for the samples that are currently selected in the food stamp QC system, then a plot of estimates from repeated samples should follow the bell-shape of the normal distribution.

The degree to which the QC system is able to provide a sound basis for accountability thus depends on the answers to two specific questions about its statistical design:

==========

Question A1: Does the present sampling and analysis design provide unbiased estimates of State error rates?

Finding: The technical design of the QC system is basically sound. Extensive empirical tests demonstrate that the estimates of overpayment-error rates are essentially unbiased estimates

---

27/The degree of sampling error could be reduced most directly by increasing the size of the annual federal and State samples. Other factors may also affect the standard error. Stratifying the caseload (drawing separate samples from categories of households) can in some cases have the effect of increasing the precision of the overall error-rate estimate (narrowing the confidence interval). However, to improve sampling precision for overall rates, stratification must entail dividing the caseload into groups, at least some of which are more homogenous (that is, have lower variance) in terms of the variable in question--the dollar amount of error--and using appropriate sampling fractions in each stratum.

of the true error rate. Moreover, the point estimate is a better measure of the true, but unknown, error rate than is the lower bound of a confidence interval. The lower bound systematically understates the true error rate.

The confidence intervals around the sample point estimates are reasonable measures of precision. The probability that the true error rate is below the computed confidence interval is somewhat lower than the nominal probability, while the probability that the true error rate is above the confidence interval is somewhat greater than the nominal probability. This modest asymmetry, however, has no effect on the unbiasedness of the point estimate and, with minor modification of current calculation methods, could be reduced to enhance the coverage of the nominal confidence interval.

Question A2: How precise are the estimates of overpayment error rates?

Finding: The estimated error rates vary from year to year within a State. Standard errors of the estimated overpayment-error rates range from a low of .004 to a high of .032, with an average of about .01 in 1985. Major improvement in precision would require much larger sample sizes.

States often stratify their samples by dividing their caseloads into segments and sampling from each segment. In some cases, samples are stratified to guarantee minimum numbers of particular types of households of interest, or simply to adjust the number of cases selected over the course of a year. In some instances, the manner in which States stratify their QC samples reduces rather than increases precision. Selecting strata more effectively and adhering to proportionate sampling (with some allowable exceptions) would improve precision.

What constitutes an "acceptable" level of sampling error depends on the cost of expanding the samples, how the estimates are used, and their consequences. The adequacy of current sample sizes and the associated precision of the error-rate estimates are policy questions discussed more fully in Chapter IV.

==========

Are Estimates of QC Error Rates Unbiased?

FNS uses a two-tier, or double, sampling scheme and a regression estimator to derive QC error rates. The point estimates derived from this procedure are estimates of the error rates that would be found if State reviews, followed by federal re-reviews, were conducted each month for every food stamp household in a State's caseload. If this approach yields more than trivially biased results, with point estimates of error rates systematically understating or overstating true error rates, it would be difficult to defend the estimates from a statistical standpoint as a basis for measuring performance and assessing liabilities for excess error rates.

Statistical theory says that, if the samples are large enough, these estimates will be essentially unbiased--regardless of the underlying distribution of errors for individual cases. However, theory does not specify how large a sample must be to be considered "large enough" to ensure unbiased or minimally biased estimates in the absence of assumptions about normal error distribution.

Thus, as part of the process of preparing this report, a study was carried out to test the accuracy and precision of the regression estimator by examining error-rate estimates based on samples from populations for which the true error rates were known.28/ Two test populations of food stamp cases were constructed to serve as simulated "State caseloads." Each test population included review cases drawn from the 1985 QC samples of four States. The test populations simulated two States that differed from each other in terms of caseload size, average benefit amount, and error rate, but which, taken together, were approximately representative of a broad range

---

28/See Morris H. Hansen and Benjamin J. Tepping, "A Statistical Evaluation of Food Stamp Quality Control," Westat, Inc. (forthcoming). The regression estimator is described in Chapter II.

of States. Repeated samples were drawn from the test populations, with each sample case replaced after it was drawn, thus greatly increasing the effective size of the test populations.29/

The study procedures consisted of drawing repeated samples from each test population to simulate State review samples, and then drawing a "federal sample" of the appropriate size from each such "State sample." The existing State review and federal re-review findings for each case in these samples were then used to compute error rates and statistics that described their relationship to the true error rates known for each test population. A total of 1,000 State samples and associated federal subsamples were drawn from each test population for each of four combinations of State and federal sample sizes. These four combinations approximate those actually used in the State and federal reviews.30/ This study in a sense simulates the equivalent of 1,000 years of drawing QC samples and estimating error rates for each test population and each sample size.

The major finding of these tests is that estimates of overpayment-error rates derived from the regression estimator are essentially unbiased. Table III.1 shows that the average of 1,000 independent samples, drawn separately for each of four illustrative sample sizes from each test population, are nearly equal to the known error rate for the population. The differences are explainable as sampling variations. Figure III.1 shows the distribution of these estimates for each of the sample size combinations from each test population. In both cases, the distribution of error rate estimates approximates very closely the desired bell-shaped curve of the normal distribution and is centered over a point very close to the true error rate. In other words, samples of the size used in the QC process, when drawn repeatedly, do in fact yield a distribution of error-rate estimates whose mean is equal to or very close to the value of the error rate that would be obtained if the entire caseload were reviewed. Thus, for sample sizes now in use, the error-rate estimates are essentially unbiased despite the skewed distribution of error amounts in the populations from which the samples are drawn.

29/The same effect could have been achieved, at greater cost, by increasing the number of cases in the test populations sufficiently to make them approximately equal in size to the various State caseloads.

30/The State and federal sample sizes used were 500/200, 800/300, 1,200/400, and 2,400/400.

Table of Contents

TABLE III.1

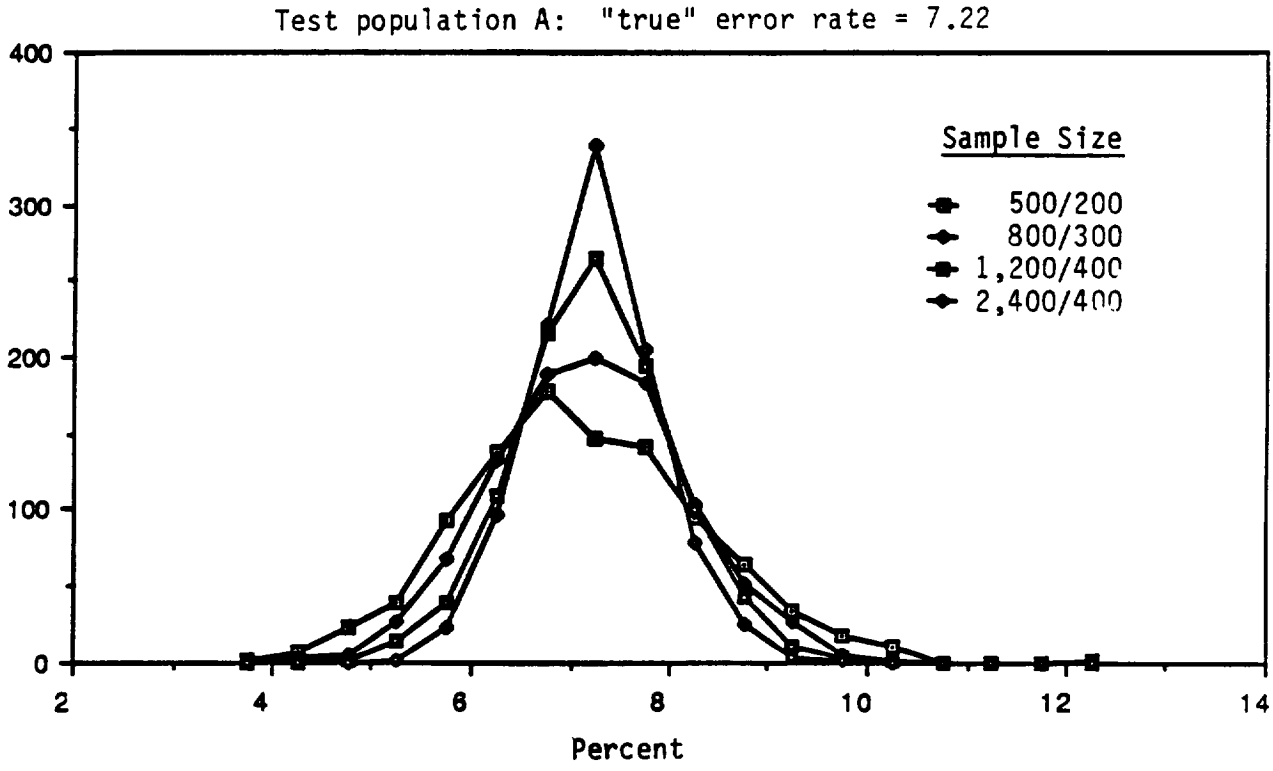
COMPARISON OF ESTIMATED ERROR RATE WITH
TRUE VALUES FOR TWO TEST POPULATIONS

Sample Size (State/federal)	Population A	Population B
True Value (Full test population)	.0722	.0825
Point estimate:		
2,400/400	.0723	.0822
1,200/400	.0723	.0826
800/300	.0718	.0827
500/200	.0714	.0822
Lower bound (90%)		
2,400/400	.0620	.0720
1,200/400	.0593	.0697
800/300	.0563	.0673
500/200	.0520	.0631

Table of Contents

FIGURE III.1

ESTIMATES OF OVERPAYMENT ERROR RATES FOR
TWO TEST POPULATIONS



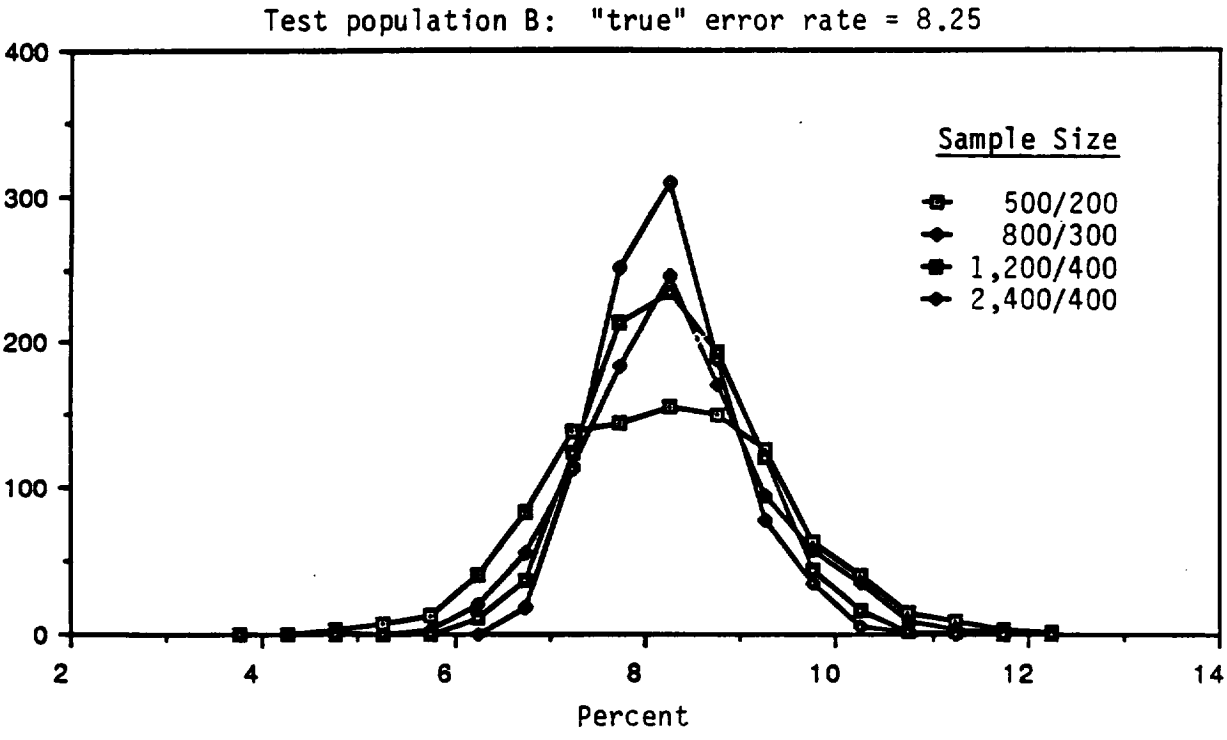


Table III.1 also shows the average of 1,000 calculations of the lower bound of a 90 percent confidence interval for each sample size and test population. In every case, the average lower bound is substantially less than the known error rate. Thus, the lower confidence bound systematically understates the true error rate.

In addition, these tests indicate that the confidence intervals as currently computed approximate reasonably closely the confidence intervals that would be found if error amounts exhibited a normal distribution. Table III.2 shows the percentage of times that the true error rate for the population was below, within, or above the computed confidence interval in the 1,000 samples from each test population. As can be seen in the table, the lower bound of the 95 percent (two-tailed) confidence interval fell above the true error rate less than, but reasonably close to, the expected 2.5 percent of the time. The upper bound fell below the true error rate somewhat more than the expected 2.5 percent of the time. A similar pattern holds true for 90 percent confidence intervals, for which the expected value in each tail is 5 percent.

This moderate asymmetry of the true confidence intervals stems from the extreme skewness in the distribution of errors and error amounts in the population. It does not affect the validity of the point estimate as an unbiased measure, but it does imply that true error rates are somewhat more likely to fall above the upper bound of the nominal confidence interval than below its lower bound.31/ Statistical procedures that would generate confidence intervals that match nominal expectations even more closely could be explored.32/

It should also be noted that the design of the regression estimator yields unbiased estimates of the true error rate regardless of how closely State review findings match the federal re-review findings. The results from State reviews, as described in Chapter II, are used to improve the precision of the error rate estimated with the regression estimator; the higher the correlation between State and federal findings, the

---

31/This, in turn, means that liability decisions based on the lower bound of the confidence interval rather than the point estimate would tend to "favor" States slightly more than implied by the nominal confidence levels.

32/Such procedures involve drawing many overlapping "replicate" samples from each sample, and estimating error rates for each replicate.

Table of Contents

TABLE III.2

CONFIDENCE INTERVAL COVERAGE FOR ERROR RATE ESTIMATES FROM TWO TEST POPULATIONS

Sample Size (State/federal)	Coverage	95% Confidence Interval[a] Population A	95% Confidence Interval[a] Population B	90% Confidence Interval[a] Population A	90% Confidence Interval[a] Population B
500/200					
	Below	2.1%	2.3%	4.4%	4.5%
	Within	94.6%	94.7%	90.0%	89.9%
	Above	3.3%	3.0%	5.6%	5.6%
800/300					
	Below	2.5%	2.1%	5.1%	4.4%
	Within	94.8%	94.3%	89.5%	89.8%
	Above	2.7%	3.6%	5.4%	5.8%
1,200/400					
	Below	2.3%	1.6%	5.0%	4.6%
	Within	95.1%	95.8%	89.3%	90.1%
	Above	2.6%	2.6%	5.7%	5.3%
2,400/400					
	Below	2.6%	1.9%	4.5%	3.7%
	Within	94.0%	93.3%	90.2%	88.9%
	Above	3.4%	4.8%	5.3%	7.4%

[a]Confidence intervals are two-tailed.

more precise the final error rate will be (i.e., the narrower the confidence interval around the point estimate). Differences between State and federal findings do not, however, introduce any bias into the final error-rate estimate. The unbiased nature of error-rate estimates does not depend on whether a particular case error is found by federal or State reviewers, or both.

The results clearly indicate that the regression estimator method is unbiased, in the sense that the average of error-rate estimates from repeated samples will be very close to the true error rate. However, error-rate estimates are based each year on a single sample, not on the average of many samples as simulated in the analysis reported herein. Thus, the remaining question is whether sample point estimates of the overpayment-error rate are sufficiently precise to serve as the basis for establishing liabilities. This question is considered in the following section.

==========

CONCLUSIONS

- Current statistical methods yield unbiased estimates of the overpayment-error rate. The expected value, or long-run average, of error-rate estimates is extremely close to true error rates. The distribution of errors in the case population does not bias the point estimate of overpayment-error rates.

- Confidence intervals as currently computed are reasonable approximations of the intervals that would be found if the distribution of the error amount were normal.

- The point estimate reflects the best available measure of a State's performance. A lower confidence bound systematically understates the true error rate (i.e., the error rate that would be found if State reviews and then federal re-reviews were conducted for the entire caseload).

==========

Are Error-Rate Estimates Reasonably Precise?

Basing estimates of such important variables as the QC error rate on samples represents a decision to avoid the much higher cost that would be associated with analyzing every case in the caseload, and to accept the "cost" associated with some degree of sampling error. It is important to understand, however, the extent to which sampling error can yield estimates that

are substantially above or below the true error rate. As noted earlier, the precision of an estimate is commonly stated in terms of its standard error, or, alternatively, as a confidence interval. This section discusses the size of the standard errors associated with current QC sampling and estimation procedures.

Table III.3 shows the State and federal sample sizes, the official estimates of the QC error rates and their estimated standard errors, and the 90 percent (two-tailed) confidence intervals computed for each State in 1985. The 90 percent two-tailed confidence interval implies that the true error rate will be above the lower bound 95 percent of the time. The standard errors range from a low of .004 to a high of .032, with an average of about .01. The table shows that the point estimates for 48 states were above the 5 percent target in 1985; for 6 of these 48 States, the confidence interval extended below the 5 percent threshold. The point estimates for 5 States were below the target; for 2 of these States, the upper bound of the confidence interval was above 5 percent.

Because the estimated error rates are based on sample results, some chance exists that the observed rate will be above the target even if the true error rate is below. Conversely, some chance also exists that the observed rate will be below the target even if the true rate is above. The approximate probability that a State whose true error rate is less than 5 percent will have an estimated error rate above 5 percent is illustrated by the following (assuming a standard error of one percent):

True Error Rate	Chance of Estimate Being above the 5% Target
5%	50%
4%	32%
3%	2%
2%	0.1%

The approximate probability that the estimated rate is below 5 percent given a true error rate above the target is illustrated by the following (again assuming a standard error of one percent):

Table of Contents

TABLE III.3

FY 1985 QC SAMPLES, PAYMENT ERROR RATES, AND CONFIDENCE INTERVALS

State	State Sample Size	Federal Sample Size	Official Error Rate (Point Estimate)	Standard Error	Confidence Interval (90%) [a] Lower Bound	Upper Bound
Alabama	2,364	400	.1350	.0095	.1194	.1506
Alaska	321	165	.1353	.0318	.0830	.1876
Arizona	2,433	400	.0938	.0074	.0816	.1060
Arkansas	1,318	430	.0788	.0110	.0607	.0969
California	2,300	413	.0708	.0073	.0588	.0828
Colorado	1,342	421	.0848	.0095	.0692	.1004
Connecticut	1,128	375	.0704	.0083	.0567	.0841
Delaware	375	178	.0717	.0135	.0495	.0939
Dist. of Col.	615	247	.0981	.0112	.0797	.1165
Florida	2,439	400	.0671	.0071	.0554	.0788
Georgia	1,214	400	.1291	.0112	.1107	.1475
Guam	301	165	.0533	.0137	.0308	.0758
Hawaii	825	306	.0435	.0081	.0302	.0568
Idaho	942	290	.0516	.0135	.0294	.0738
Illinois	2,136	400	.0816	.0069	.0702	.0930
Indiana	1,314	423	.1090	.0081	.0957	.1223
Iowa	1,120	408	.0841	.0086	.0700	.0982
Kansas	1,026	361	.0816	.0125	.0610	.1022
Kentucky	1,595	400	.0600	.0051	.0516	.0684
Louisiana	1,230	406	.0976	.0092	.0825	.1127
Maine	1,060	353	.0791	.0093	.0638	.0944
Maryland	1,194	403	.0737	.0043	.0666	.0808
Massachusetts	1,207	415	.0971	.0102	.0803	.1139
Michigan	1,734	402	.0735	.0078	.0607	.0863
Minnesota	1,218	404	.0951	.0144	.0714	.1188
Mississippi	1,259	400	.0798	.0098	.0637	.0959
Missouri	2,413	400	.0523	.0042	.0454	.0592
Montana	780	284	.0744	.0064	.0639	.0849
Nebraska	1,301	401	.0904	.0097	.0744	.1064
Nevada	541	222	.0248	.0045	.0174	.0322

TABLE III.3 (continued)

State	State Sample Size	Federal Sample Size	Official Error Rate (Point Estimate)	Standard Error	Confidence Interval (90%) [a] Lower Bound	Upper Bound
New Hampshire	448	203	.0442	.0089	.0296	.0588
New Jersey	2,307	428	.0850	.0062	.0748	.0952
New Mexico	1,503	427	.0883	.0062	.0781	.0985
New York	1,251	401	.0711	.0082	.0576	.0846
North Carolina	1,256	402	.0649	.0089	.0503	.0795
North Dakota	384	190	.0353	.0080	.0221	.0485
Ohio	2,390	403	.0743	.0057	.0649	.0837
Oklahoma	1,335	427	.1058	.0129	.0846	.1270
Oregon	3,203	439	.0941	.0122	.0740	.1142
Pennsylvania	1,160	396	.0936	.0119	.0740	.1132
Rhode Island	1,106	399	.0800	.0085	.0660	.0940
South Carolina	1,730	400	.1210	.0090	.1062	.1358
South Dakota	582	238	.0315	.0074	.0193	.0437
Tennessee	1,255	400	.0639	.0068	.0527	.0751
Texas	1,261	420	.1038	.0101	.0872	.1204
Utah	573	227	.0726	.0118	.0532	.0920
Vermont	607	240	.0806	.0125	.0600	.1012
Virgin Islands	302	163	.0973	.0105	.0800	.1146
Virginia	1,210	413	.0667	.0059	.0570	.0764
Washington	2,497	423	.0950	.0073	.0830	.1070
West Virginia	1,388	405	.0507	.0051	.0423	.0591
Wisconsin	2,898	407	.0800	.0083	.0663	.0937
Wyoming	326	165	.0678	.0146	.0438	.0918
Total	70,115	18,688	.0830			

[a]Two-tailed 90% confidence interval; approximately 95% of the time, the lower confidence bound will be below the true error rate.

True Error Rate	Chance of Estimate Being below the 5% Target
5%	50%
6%	32%
7%	2%
8%	0.1%

As long as quality control reviews are performed for samples of a State's food stamp caseload rather than for all cases, some degree of sampling error is inevitable. Given the two-stage review process and the regression estimator, the degree of sampling error could be reduced most directly by increasing the size of the annual federal and, also, the State samples when the federal sample is a large fraction of the State samples.33/ An alternative way to increase the effective sample size is to accumulate sample results and liability calculations over a longer period of time.

Larger samples are obviously more expensive. While precise figures are not available, FNS estimates that the current federal re-review process costs about $200 per review. Thus, doubling the federal sample size (from 400 cases per year in the larger States to 800) would cost about $4 million per year. Similarly, each State review costs about $475. If all States were required to review at least 2,400 cases (the current maximum in the larger States), the additional State and federal costs would be over $25 million.

Given any particular sample size, however, the precision of error-rate estimates can also be affected by the use of sample stratification. Stratifying a sample entails dividing the annual food stamp caseload into strata, or segments, and sampling separately from each stratum. States may choose to stratify their samples for several reasons: to maximize the overlap between AFDC and food stamp QC samples, to achieve the

---

33/The standard error of the regression estimate can also be reduced less directly by increasing the correlation between federal and State review findings for cases in the federal sample. However, because there are no clear-cut ways to do so, this approach is not discussed.

desired representation of various geographic regions within a State, or to account for changes in the sampling rate.34/

In preparing this report, FNS reviewed State stratification methods currently in use to evaluate the general effect of stratification on the precision of error-rate estimates.35/ The analysis focused on the extent to which the sample stratification now in use affects the variances (i.e., the precision) of the estimated error rates, given the overall sample size constraint. The study was carried out using State QC data for Fiscal Years 1983-85 for States that selected stratified samples, and in which sampling rates varied across strata (i.e., in which the State did not follow strictly proportionate sampling). Based on the 64 samples from these States for these three years, the variance of the error-rate estimates was computed using the stratified samples as actually drawn and the regression estimator as now used. The variance was then compared with variances that would result from using several alternative methods for stratifying the sample and calculating error rates from the stratified samples.36/

The primary conclusion of this analysis is that moderate losses in precision (i.e., larger standard errors) result from using stratification with variable--as opposed to proportioning--sampling fractions.

34/Some States adjust the sampling fraction in later months of the year as actual caseloads depart from projections, to ensure a total annual sample of the desired size. These States adjust sampling fractions to avoid drawing a review sample larger or smaller than is required to achieve the target sample of completed reviews; if the sampling ratio were left constant, the State might have to spend more staff time on reviews to complete the larger sample, or accept an upward adjustment in its error rate if the number of reviews falls short of the required minimum.

35/Based on its audit of the 1984 QC process, the U.S. General Accounting Office found that in one State the small size of some sample strata affected the statistical validity of the State's error-rate estimate. GAO recommended that FNS set a minimum standard for stratum size to prevent such problems. The analysis undertaken by FNS considered this GAO recommendation (see footnote 37).

36/Again see Hansen and Tepping, "A Statistical Evaluation of Food Stamp Quality Control."

For 14 of the 64 stratified samples studied, stratification led to a larger variance of the error-rate estimate than would occur with a simple random sample of the same size. Better precision can be achieved by proportionate or approximately proportionate sampling from strata. As a general rule, adhering to proportionate sampling at a uniform rate across months is also preferable. Assuming that caseload projections are reasonably accurate, so that the resulting samples do not differ sharply from the target sample size, proportionate sampling will yield better precision per unit of cost even if the number of completed reviews falls slightly short of or above the minimum required sample. If States must reduce sampling fractions late in the year because of budget constraints, however, modest changes in the sampling rate late in the year can be allowed without seriously increasing the variance of the error-rate estimates.37/

Taking sampling error into account is important because of the uncertainty it injects into the decision to hold States liable for excessive error rates. Even though the point estimate is unbiased, there is some risk that the liability assessed in any given year would be substantially higher or lower than the liability that would be assessed if the true error rate were known.

What constitutes an "acceptable" level of sampling error, however, depends on the cost of expanding samples to obtain more precise estimates, how the estimated error rates are used, and their consequences--in this case, how error rates are compared with performance standards, and how the costs of errors (which are now borne primarily by the federal government) are shared with States when there is "excess" error. In the end, therefore, the adequacy of the current sample size and the associated precision of error-rate estimates is a policy question rather than a technical one. The implications of these findings for the current liability process are discussed more fully in Chapter IV.

37/The implication of this analysis is that setting a minimum sample size for all strata, as recommended by the GAO, should be avoided, because it most likely implies a departure from proportionate sampling.

CONCLUSIONS

- Although the error-rate estimates are unbiased, estimated error rates will vary from sample to sample within a State simply because of sampling errors. Using the point estimate will sometimes yield an estimate that is too high or too low for a State in any single review period. Under current procedures, however, these risks are shared by State and federal governments.

- The acceptable amount of sampling error in estimates of error rates used for liability decisions is essentially a policy decision. The primary way to reduce sampling error is to increase the sample sizes. Thus, a tradeoff exists between desirable improvements in precision and the cost of increasing sample sizes.

- Current procedures for estimating error rates from stratified samples are basically sound. However, some improvement in the precision of error-rate estimates per unit of cost could be achieved by adhering to proportionate or nearly proportionate sampling from sample strata, even if adhering to proportionate sampling leads to modest reductions in the number of sample cases.

B. ACCURACY AND RELIABILITY OF REVIEW OPERATIONS

The results described in Section A indicate that the technical design of the QC system is basically sound and can be expected to yield unbiased estimates of overpayment-error rates, with a degree of precision that can be influenced by tradeoffs between desired precision and the costs of reviews. Even with a technically sound design, however, the integrity of QC results depends on the accuracy and reliability with which the design is implemented in State reviews, federal re-reviews, and error-rate calculations. Thus, a second important component of this assessment of the QC system focuses on how the system is operated. Four specific questions are examined:

==========

Question B1: Is QC policy properly and accurately followed in State and federal reviews of active cases and error rate calculations?

Finding: Repeated audits by the General Accounting Office and the USDA Office of Inspector General confirm a generally high level of conformance to federal requirements and standard practice in the application of QC procedures by State and federal reviews. FNS has centralized the calculations of the regression estimates to ensure that statistical methods are applied consistently for all States.

Question B2: Are adequate controls and incentives in place to ensure that reviews of sampled cases are completed?

Finding: FNS has made several modifications to QC procedures in order to prevent the unwarranted incompletion of reviews, and several further changes are being considered.

Error-rate adjustments based on the rate of incomplete reviews have a small marginal effect on final error rates and, in general, are a reasonable method for strengthening incentives for completing reviews. These adjustments, however, produce a marginally greater penalty for States with small samples and low overpayment rates.

The current method for adjusting error rates based on incomplete reviews creates a strong incentive for some States to adjust sampling fractions late in the year, a form of disproportionate sampling. As described in Section A, this practice can reduce the precision of error-rate estimates.

Question B3: Do differences in State QC approaches or review costs--the "level of effort" devoted to QC--create any systematic differences in the error-rate results?

Finding: An analysis of State QC expenditures and reported error rates shows no significant correlation, and thus no evidence that some States have higher error rates because of more careful review procedures. Moreover, the two-stage review process is designed to compensate for potential differences across States. One important role of the federal subsample is to help ensure that the food stamp QC system is administered with reasonable uniformity and comparability.

Question B4: In what way does the current two-stage process of State and federal reviews affect the overall operational accuracy and cost-effectiveness of the QC system?

Finding: As reported in Section A, current statistical methods for using findings from both State and federal reviews yield unbiased estimates and increase the precision of error-rate estimates by drawing on data from both samples, relative to using only the results of the federal sample. The two-stage process strengthens the statistical soundness of error-rate results.

Some questions remain, however, about whether a two-stage review process leads to higher overall costs than would be incurred in a single-stage review. A current FNS demonstration is testing the feasibility and cost of a single-tier federal review process, evaluating the difference in review findings, and testing a standard review protocol. This demonstration will yield

further information that may be used to improve the effectiveness of the QC process.

==========

Are Reviews Undertaken Accurately and in Accordance with Policy?

The quality of the performance measurement data generated by the QC system rests first of all on the ability of State and federal QC units to draw samples according to established procedures and to conduct case reviews accurately--that is, their ability to distinguish between correct and erroneous cases. This issue could not be assessed independently during the course of the current study because of its limited timeframe. Instead, FNS relied on an indirect test to address the question--namely, whether previous independent reviews and audits showed any evidence that major problems exist. The accuracy of QC operations has been examined extensively in audits over the past several years by two independent review bodies: the U.S. General Accounting Office (GAO) and the Office of the Inspector General of the U.S. Department of Agriculture.38/ The overall results of these audits support the conclusion that sample selection and case reviews generally conform with federal requirements and standard practice. Minor problems found in these audits have not had any substantial effects on either error rates or State liabilities.

Two issues raised in GAO audits with respect to the inconsistent use of statistical procedures across States have been addressed by FNS. FNS now carries out all of the regression estimate calculations centrally for all States, at its Washington Computer Center, using a standard computer program, thus guarding against mathematical errors or inconsistencies across regions. FNS has also standardized procedures for weighting results across sample strata for States that use stratified samples; all weighting is now based on caseload size in each stratum.

38/See U.S. General Accounting Office, "Food Stamp Program: Refinements Needed to Improve Accuracy of Quality Control Error Rates," September 1986; GAO, "Food Stamp Program: Statistical Validity of Agriculture's Payment Error-Rate Estimates"; and Office of Inspector General, U.S. Department of Agriculture, "Nationwide Audit of the Quality Control Error Rate Reduction System," October 1986.

CONCLUSIONS

- The implementation of QC sample selection and case review procedures appears to conform with federal requirements and standard practice, based on audits by the GAO and Inspector General.

Are Appropriate Rules in Place to Minimize Incomplete Reviews?

Failure to complete reviews of cases selected in QC samples can, if it occurs frequently, undermine the accuracy of error-rate estimates. If the likelihood of error differs systematically between cases for which reviews are and are not completed, a high rate of incomplete reviews could introduce a "nonresponse" bias in estimates of the error rate. The present QC system contains procedural requirements to promote the completion of reviews, as well as an error-rate adjustment that creates an incentive to minimize incomplete reviews.

QC procedures are meant to encourage the completion of reviews whenever possible, although some circumstances remain in which it is impossible to complete a review.39/ In addition to procedural requirements, the current QC system creates an incentive to complete reviews by providing for an upward adjustment of a State's regressed error rate if reviews are completed for less than 100 percent of the sampled cases. FNS makes this adjustment by attributing a higher error rate to

---

39/A recent GAO audit recommended changes in QC procedures to expand the universe of cases subject to review and to relax requirements for completing case reviews. The recommendations proposed (1) requiring applicants to sign information-release authorizations, (2) broadening the circumstances in which reviews can be completed without an interview, (3) counting errors where reasons for ineligibility have been found but formal review procedures cannot be completed, and (4) completing reviews for all cases under fraud investigation. These recommendations raise questions about their possible conflict with State privacy laws and potentially erroneous assumptions of error; they remain under study by FNS. The GAO recommendations can be found in "Food Stamp Program: Refinements Needed to Improve Accuracy of Quality Control Error Rates," U.S. General Accounting Office, September 1986.

incomplete review cases than is found among completed cases. The attributed rate is set at a rate two standard deviations above the overall computed error rate.40/ Under this procedure, the adjustment is based on the standard deviation of the State's reported error rate.41/

FNS has evaluated the details of these features as part of this study. Three issues were examined:

1. The effect of the incomplete review adjustment across States

2. The appropriate base for determining the incomplete review adjustment

3. The possible use of differential completion rates across sample strata to calculate error rates

It is important first, however, to place the issue of incomplete reviews in the context of their relatively minor effect on measured error rates. States complete a very high percentage of review cases. In 1985, States completed reviews for an average of 98 percent of their sample cases; only five States had completion rates below 95 percent, and none less than 92 percent. Given the present method for calculating adjustments based on incomplete reviews, their effect on error

---

40/The formula for the adjusted rate is:

$$ER_{adj} = (1 - c)r + c(r + 2s_r) = r(1 + 2cs_r/r),$$

where r is the regression estimate of the payment-error rate, c is the proportion of reviews not completed, and s_r is the standard error of the payment-error rate as estimated from the State sample. The adjustment factor is $(1 + 2cs_r/r)$.

41/The GAO has suggested that FNS use the standard deviation of the regressed error rate rather than the error rate from the State sample, because the regressed error rate is based on more information. In the view of FNS, this issue is a minor policy issue rather than a statistical one. The current adjustment is based on a somewhat arbitrary policy definition; there is no clear statistical basis for choosing one or the other standard deviation as the adjustment basis.

rates is very small. This effect depends on the completion rate, the regressed error rate, and the standard error of the State-reported error rate. Given the plausible range of values for these last two factors, the adjustment for a State with a 95 percent completion rate, for example, could range from 1.005 to 1.03.42/ Such a State with a regressed error rate of 6.50 percent, for instance, would thus end up with an official error rate of between 6.53 and 6.69.

The Effect of the Incomplete Review Adjustment. FNS examined the "neutrality" of the adjustment for incomplete reviews.43/ As can be seen in the formula for the adjusted error rate, (see footnote 17), the greater the standard error of the sample estimate (s_r), and the lower the estimated error rate (r), the higher the adjustment factor. Given a certain rate of incomplete reviews, this formula produces greater upward adjustments for States with small samples and States with low overpayment rates, because each of these, on average, is associated with an increase in the value of the last term in the adjustment formula.

In reality, this issue is relatively minor, because the adjustment for incomplete reviews makes very little difference in the final error rate. This bias in the incomplete review adjustment, which works against States with small samples or low error rates, can be eliminated with a slight modification of the formula for the adjustment factor. In lieu of the term s_r/r in the adjustment factor, a constant value could be substituted. A constant value of .1 or .2 would yield approximately the same adjustment levels as the current formula, but the adjustment factor would not be sensitive to sample size or the error rate.

The Base for Computing the Incomplete Review Adjustment. As pointed out earlier, maximizing the precision of error-rate estimates from QC samples requires a nearly constant sampling rate throughout the year. The current procedure for adjusting error rates based on incomplete reviews creates an incentive for States to increase their sampling rates late in the year if their actual caseload is lower than projected. They do so

---

42/The adjustment factor is the value by which the regressed error rate is multiplied to arrive at the final adjusted error rate.

43/See Morris H. Hansen and Benjamin J. Tepping, "A Statistical Evaluation of Food Stamp Quality Control," Westat, Inc. (forthcoming).

because the completion rate is based on the number of cases selected that are subject to review or the minimum required sample size, whichever is higher. Such increases in sampling rate can reduce the precision of error-rate estimates more than modest reductions in sample size in some circumstances. Although increases (or decreases) in the sampling rate in the last few months of a year would not substantially reduce the precision of error-rate estimates, it is preferable to maintain a uniform sampling rate throughout most of the year.

Differential Completion Rates Across Sample Strata. FNS has examined whether data from stratified samples should be adjusted if differential review completion rates exist across strata. Under current procedures, the results found in sample strata are weighted by the size of the caseload subject to review in each stratum to arrive at an overall error rate. An alternative would be to adjust the caseload figure used to weight error data based on the completion rate for each stratum. This approach would give greater representation in the calculation to strata with high completion rates; in effect, it would be an attempt to represent completed review cases rather than the total review sample.

The conclusion of FNS is that the appropriate regressed error rate is one which provides a relatively precise estimate of the error rate for the caseload at large, rather than giving higher weight to completed cases. In stratified samples, the estimate of the overall error rate must be derived from sound estimates of the error rates for each stratum. Although some attention might properly be paid to the method for imputing data to missing cases, that issue arises in the context of attempts to obtain an estimate of the error rate for the full population. Thus, current procedures for weighting data from sample strata without adjusting for strata completion rates constitute a sound approach.

==========

CONCLUSIONS

- Current procedures that make an upward adjustment to a State's error rates if it fails to complete reviews on 100 percent of its sample cases are reasonable. The effect of the adjustment on the final error rates is marginal.

- The formula for adjusting regressed error rates affects States more heavily if they have small samples or low error rates. The current procedure for adjusting error rates for incomplete reviews also creates some incentive to increase

sampling rates when caseloads are below original projections. Sharp variation in sampling rates can reduce the precision of error-rate estimates.

==========

Does Greater State "QC Effort" Lead to Higher Error Rates?

Although federal QC handbooks prescribe QC procedures for all States, some variations in State QC practices are inevitable, and in some cases desirable. Staffing levels, the quality and intensity of supervision, and the degree to which reviewers' case results are reviewed before being submitted to FNS are elements of the QC process which may vary across States. Arguments can be advanced that these variations in the review practice can contribute to differences in measured error rates independently of the effect of true differences in program performance.

On the one hand, greater staff resources, and better supervision of reviewers, might mean that reviewers conduct more thorough reviews and uncover more errors. On the other hand, the intensity of the review effort may reflect a strong commitment of State managers to the QC process. If the information produced by QC is fed back to program managers, it may be an important factor in improving performance and lowering error rates. Alternatively, a more systematic and painstaking reexamination of reviewers' results by State QC oversight staff can discover policy interpretations which overturn reviewers' error findings before they are submitted to FNS. Some States conduct such reexaminations of reviewers' results only for error cases, while others do so for both error and correct cases. It is possible that such differences could introduce marginal differences in reported error rates.

As pointed out earlier, however, findings from the State review serve primarily to improve the precision of the error-rate estimates derived from the federal re-review sample. A State's ability to find more or fewer errors primarily affects the standard error of the estimated error rate. One of the functions of the federal re-review is to find errors that the States do not. If the federal reviews are successful, differences in State-reported errors do not bias the regressed error rate.

While no conclusive evidence on the variation in the ability of States to find errors can be offered, some recent analysis failed to find any clear connection between available measures of State QC effort and error rates. An exploratory analysis of the relationship between State payment-error rates and two measures of the level of QC effort expended by each State

found a correlation of only .16 between State-reported error rates and the State costs of a QC review, and a correlation of -.25 between a State's error rate and the number of reviews completed per staff member.44/

==========

CONCLUSIONS

- There is no conclusive evidence that variations in State review practices have a systematic effect on error rates, although such effects may occur. The QC system is designed so that findings from the State review serve to improve the precision of the error-rate estimate derived from the federal re-review sample. A State's ability to find more or fewer errors affects the precision of the error rate, but not the point estimate.

==========

Does the Two-Stage Review Affect Accuracy and Cost?

A careful analysis of present sampling methods and procedures for calculating error rates from sample data, as reported in the preceding sections, indicates that the technical design of the QC system is basically sound. These procedures have been developed to increase the accuracy of error-rate estimates given a set of State and federal review roles that have developed over time. Although the two-stage process that involves both State and federal reviewers has been found in this study to be technically sound, the question can also be raised about whether it is the best approach for generating accurate error-rate measures. Two specific concerns merit examination:

1. Would error rates based on a single federal review differ from error rates obtained with a State review followed by a federal re-review?

44/It must be acknowledged, however, that these results are very tentative, because of (1) considerable variation in the manner in which States measure their QC costs, and (2) the wide range of factors that can affect QC costs and the productivity of reviewers that do not necessarily have any consistent relationship to the actual time devoted to case reviews.

2. Would it be more cost-effective to base error rates simply on a federal review sample (possibly expanded) and to allow States to establish their own review procedures for management information purposes?

A consideration of major structural changes to the QC process is beyond the scope of this study, and beyond the scope of analysis performed by FNS for this report. However, FNS has initiated a longer-term investigation of these issues by sponsoring a pilot demonstration and evaluation of a one-tier QC process. The pilot projects, in North Carolina and Missouri, were implemented in October 1986 and will run for a full year; evaluation results are expected in 1988.

In the one-tier QC demonstration, an entirely federal review process is running parallel to the current two-stage review process, and the results and costs for the two systems will be compared. A structured worksheet has also been developed to provide a consistent and comprehensive guide to the review process. The evaluation of the demonstration will address three questions associated with the two issues pointed out above:

1. Is a one-tier federal process administratively feasible?

2. Is a one-tier system less expensive than the present two-stage review process?

3. What is the nature and frequency of State disagreements with federal error findings?

This evaluation will enhance an understanding of the advantages and disadvantages of changing to a one-tier system. It will not, however, address one factor in the assessment of relative costs: the extent to which State reviews--for management information purposes--would continue to be performed and constitute an additional federal cost beyond the cost of the one-tier federal review.

==========

CONCLUSIONS

- Current procedures for using review findings from two stages (State and federal) are fundamentally sound. However, more cost-effective approaches may be possible. The evaluation

of a one-tier QC demonstration will provide further information on the administrative feasibility, costs, and possible effects on error rates associated with switching to a one-tier federal review process.

==========

C. THE DEFINITION OF QC ERROR: WHICH ERRORS SHOULD BE COUNTED?

Section A of this chapter focused on issues surrounding the statistical design of the QC system, and Section B on the operation of the system. This section deals with issues associated with the definition of the error rate--which types of errors should be included in the basic performance measure used as a basis for judging State performance. The appropriate error definition is largely a policy choice. The issue is important because the types of errors included in the error rate affect the types of incentives for program administration created by the QC system. A change in definition can also change relative State performance. The policy decision can be informed by the operational experience of USDA and the States. Four specific questions are examined in this section:

==========

Question C1: Should the error rate include only "agency errors" and exclude "client errors," or should it continue to include both types of error? Stated another way, should States be held accountable for client errors?

Finding: The distinction between agency and client errors is a difficult and often ambiguous one; classification depends in many instances on the judgment of individual reviewers. Excluding client errors from the official error rate could encourage the classification of more errors as client-caused. The relatively subjective classification of errors could bias error rates across States to an unknown and unmeasurable degree.

Holding States accountable for client errors maintains balanced incentives. Many client errors are in fact problems that can be addressed by agencies (with collateral contacts or computer matches, for example) and should be included in the error-rate measure.

Question C2: Does the present definition of QC errors include errors that are strictly procedural and do not affect the issuance amount?

Finding: Strictly procedural errors are not included in the food stamp error rate.

Question C3: Does the error-rate measure capture the net cost of error to the federal government for households that participate in multiple programs?

Finding: Current procedures accurately measure the loss to the Food Stamp Program based on the actual income available to each household. An error measure designed to capture the net cost to the federal government could create food stamp errors where none exist and increase administrative complexity and burden.

Question C4: Does the exclusion of underpayment and negative action errors from the basis for fiscal liabilities under current law create an imbalance in the incentives to reduce the incidence of all errors?

Finding: There is no evidence that underpayment error rates have increased as States reduced overpayment-error rates. The evidence to date suggests that, in general, over- and underpayment rates generally move together, suggesting that administrative performance affects both positive and negative errors similarly. Underpayment rates have been relatively stable or have fallen slightly in recent years.

Underissuances are measured as part of the same sampling and review process as overpayment errors, and are measured in the same units. A measure of the "cost" of underpayment could be incorporated in the official error rate. Doing so could improve the balance of error-avoidance incentives by clearly and formally stating that the QC system does not tolerate underissuances.

Negative action reviews are a procedural review based on a separate and quite different sample. These reviews yield an estimate of the number of cases that were possibly eligible for benefits but involved a procedural deficiency in their denial or termination. Thus, negative action reviews do not provide a payment-error rate comparable to the rate available for over- and underissuances. Whereas error rates derived from active household samples present a clear statement about the frequency of incorrect issuance, negative action review results do not.

Although the negative action review provides important guidance on problems in the certification process, combining it in a single error-rate measure with the results of active case reviews would require major changes in the QC process.

==========

Should Client Errors Be Excluded from the Error Rate Measure?

In the current QC review process, reviewers assign a code to each error to distinguish between "client errors" and "agency errors." State agencies have at times argued that errors caused by client action or inaction should not be incorporated into the error rate that serves as the basis for fiscal liabilities: the QC error measure should capture only agency errors. This argument is based on the view that client errors are beyond the agency's control. The resolution of this issue is important because the types of errors counted and not counted in the error rate can affect the direction and intensity of State efforts to reduce the incidence of errors.

The distinction between client and agency errors is not always clear and unambiguous. For example, when clients fail to report income, it may be that case workers have not adequately explained the reporting requirements to clients. Second, the client's responsibility to report information must be balanced against the agency's responsibility to follow up on questionable, incomplete, or inconsistent information provided by clients. For example, if a 70-year-old male applicant reports no Social Security income, and a QC reviewer later detects the unreported income, both parties were at fault. Determining the responsible party is difficult. Errors classified as client errors by one reviewer, or in one State, can often

justifiably be classified as agency errors by another, and vice versa.45/

Client errors, moreover, are not entirely beyond the control of State agencies. Agencies can and do take a variety of steps to prevent client errors, and generally try to structure the interaction between clients and staff in such a way as to reduce client error. These measures include such steps as improving interviewing methods, strengthening verification procedures, and conducting computer file matches. A recent FNS regional review of one State's QC findings hints at the effectiveness of such measures. Of the 24 cases with earned-income errors in the State sample, FNS found that computer wage matching would have detected 20 of the errors, although the State held the client responsible for 14 of the cases. Client errors can be influenced by managing the interaction between staff and clients.

==========

CONCLUSIONS

- The QC system under current law encourages the reduction of all types of errors. The distinction between agency- and client-caused errors is ambiguous. The relatively subjective classification of errors could bias the measure in unknown ways and to an unmeasurable degree.

==========

Does the Error Rate Include Purely Procedural Errors?

The emphasis of the food stamp QC system is on measuring the payment-error rate, based on errors that lead to the incorrect issuance of food stamp benefits to households. The Food Stamp Act requires that certain procedures be followed and case documentation established in addition to the applica-

---

45/Some indication of the degree of interpretation involved in distinguishing between client and agency errors is provided by the variation across States in terms of the relative number of errors in each category. The proportion of variances attributed to clients by the States ranged from 31 percent to 75 percent in Fiscal Year 1985. Seven States attributed less than 40 percent of the variances to clients; nine attributed more than 60 percent of the errors to clients. (Variances refer to error elements, not to the dollar amounts of errors.)

tion of basic eligibility requirements in the certification process. However, the present food stamp QC system focuses on measuring a direct outcome: the correctness of benefit payments. Errors are counted only if correcting the problem would have directly changed the eligibility status of a household or its benefit amount.

Thus, QC errors in the Food Stamp Program consist of errors involving eligibility rules that pertain to financial circumstances, household composition, or individual characteristics. Beyond these basic eligibility rules, QC errors are counted when recipients refuse to meet three program requirements that do not pertain to household circumstances or characteristics: the requirement that certain individuals register for work, the requirement that all participants provide a Social Security number, and the requirement that certain households file a monthly report. If an individual actively refuses to comply with these rules and no action is taken to terminate or reduce benefits, a QC error is counted. In all of these cases, the committed error directly contradicts an eligibility requirement, and correcting the error would affect benefits.

Strictly procedural errors that have no direct effect on benefits, however, are not counted in the food stamp error rate. Such procedural deficiencies could include:

- Failure to have the application signed
- Agency failure to register a household member for work
- Agency failure to confirm that all members have Social Security numbers or are applying for them

In these cases, the errors committed do not necessarily lead to an incorrect benefit. Such errors could indirectly contribute to issuance errors, since they might be indicative of systematic problems in administering the program. If this is the case, the QC system will capture those errors where they matter--that is, when benefits are incorrect. Correcting the procedural error itself would not alter the eligibility or benefits of the household. Excluding purely procedural errors from the error rate is consistent with the intent of the food stamp QC system to measure the rate of erroneous issuance.

CONCLUSIONS

- Current QC procedures focus on substantive errors that have a direct effect on eligibility status or benefit amount.

Table of Contents

loss to the federal government from an error that affects both programs.46/

The effect of the alternative procedure on the calculated error amount can be illustrated for a household that receives both AFDC and food stamps and fails to report other unearned income, as shown in Table III.4. Suppose that a household which receives both AFDC and food stamps had $100 in unreported monthly income--in this example, a total of $200 in income rather than the $100 reported at certification. If the household had reported the income, AFDC benefits would have been $100 lower, and the AFDC QC process would record a $100 overpayment. Under present food stamp QC procedures, the food stamp error calculation will focus on the $100 in unreported income and the original $200 AFDC benefit because it was the AFDC income actually received by and available to the household that month. Food stamp QC reviewers would conclude that total net income was $100 higher than reported, and based on the 30 percent benefit reduction rate in the Food Stamp Program would compute the food stamp overissuance as $30. Total QC errors in both programs would be $130.

If food stamp QC procedures were defined to measure the "net federal loss" of a food stamp error beyond the cost to other programs, the dollar value of the food stamp error would come out differently. Under this approach, food stamp QC reviewers would conclude that net income in the absence of the reporting error would have been the same as originally determined, because the increase in unearned income would be offset by a decrease in the AFDC grant. The value placed on the food stamp error would then be zero rather than $30.47/

46/Concern about the present error valuation method is described in Center on Budget and Policy Priorities, "Does the Food Stamp Error Rate Overstate the Loss to the Federal Government through Errors?" June 1985.

47/In cases of unreported earnings, the error "offset" is more complicated, because the recomputation of both the AFDC benefit and the food stamp benefit involves earned-income deductions. For unreported earnings of $100, current procedures would find an AFDC error of $67 and a food stamp error of $24, whereas the alternative would find an AFDC error of $67 and a food stamp error calculated as $3.90 (but which would not be counted at all because it is under $5).

Table of Contents

TABLE III.4

ILLUSTRATION OF DIFFERFENCE IN FOOD STAMP ERROR CALCULATIONS UNDER CURRENT PROCEDURES AND "NET FEDERAL COST" APPROACH

Factors in Certification and Error Calculation	Original Food Stamp Certification	FS Error Calculation Under: Current Procedures	FS Error Calculation Under: Net Federal Cost Approach
Uhearned Income	$100	$200	$200
AFDC	200	200	100
Total	300	400	300
Food Stamp Allotment[a]	210	180	210
AFDC Error		100	100
Food Stamp Error		30	0
Total Error		130	100

[a]The food stamp benefit calculation is for households of four (maximum allotment = $271) with no deduction other than a standard deduction of $99.

To estimate the differences in the error-rate outcomes generated by these two approaches, a simulation study was performed using the 1984 QC sample.48/ Records for food stamp households with AFDC in the QC sample were recoded to eliminate food stamp errors associated with unearned income, and to adjust the amount of errors associated with earnings and household composition. The error rates were then recomputed.49/

As shown in Table III.5, taking into account the offsetting effect of changes in AFDC benefits when food stamp errors are measured would have lowered the national 1984 food stamp payment-error rate. For the pure AFDC/food stamp households for which this simulation was performed, the reported error rate would have been reduced from 5.3 percent to 4.0 percent. The resulting effect on the overall error rate would have been a 6.3 percent reduction, from 8.6 to 8.1 percent. It should be noted, however, that this estimate does not account for the possibility (as described below) that the net federal cost approach could create new types of errors.

Both current procedures and this alternative approach have appealing justifications. Current procedures base food stamp errors on the actual circumstances that confront the food stamp eligibility worker and on the actual amount of money available to the household in the review month. The alternative approach bases food stamp errors on a measure of the incremental financial loss to the federal government when an error affects multiple programs.

Careful consideration must be given to the complexity of the administrative procedures that would be necessary to adopt the net federal cost approach, and the secondary effects that this approach could have on error-rate measurement. The drawbacks described herein emerged clearly in responses to the rule proposed by FNS to adopt a net federal cost approach in 1979, and remain relevant comments on such a proposal today.

48/See James C. Ohls and Jennifer Schore, "Potential Effects of Program Changes on Food Stamp Program Error Rates," Mathematica Policy Research, Inc. (forthcoming).

49/This simulation entailed recoding only pure AFDC/food stamp households, since, for "mixed" AFDC/food stamp households, it was generally impossible to determine whether errors involved the AFDC or non-AFDC members. Consequently, the likely effect on error rates is understated.

Table of Contents

TABLE III.5

ESTIMATED 1984 ERROR RATE REDUCTION WITH AFDC OFFSET CONSIDERED IN THE CALCULATION OF FOOD STAMP ERROR

	Household Type		
	Pure AFDC	Other	All Households
Payment Error[a]			
1984 Rate	5.3[b]	11.0[b]	8.6
Estimated Percentage Reduction	24.5%	-	6.2%
Adjusted 1984 Rate	4.0	11.0	8.1
Underpayment Error			
1984 Rate	1.7	2.7	2.3
Estimated Percentage Reduction	28.0%	-	8.5%
Adjusted 1984 Rate	1.2	2.7	2.1
Case Error			
1984 Rate	19.5	25.1	23.4
Estimated Percentage Reduction	6.0%	-	1.6%
Adjusted 1984 Rate	18.4	25.1	23.0

NOTE: Analysis is based on a weighted sample of case data from July-August 1984.

[a] Includes overissuances and issuances to ineligibles.
[b] Weighted rates are derived from the national average of regressed QC error rates.

- The net federal cost approach would create food stamp errors where none existed before. For instance, if an error had been made in computations or policy interpretations that were relevant only to AFDC, the QC reviewer might determine that the correct AFDC grant should have been higher (or lower), and thus that a food stamp error was made. A food stamp error would be counted even though food stamp policy had not been applied erroneously, and eligibility and benefits had been determined correctly.

- A consistent approach would account for adjustments not only in AFDC, but also in other benefit programs. This comprehensive approach would increase the burden on reviewers. FNS surveys show that 20 percent of food stamp households receive Social Security, 18 percent SSI, and 12 percent General Assistance (GA). Determining the correct benefit amount in those programs when income misreporting has occurred, and then adjusting the food stamp error, would substantially complicate the task of the food stamp QC reviewer.

- Adopting the net federal cost approach is feasible only if a State has integrated sampling and reviews. More than half of the food stamp agencies (28 of 53) still use separate review processes for food stamps and AFDC. Adopting the alternative procedure would force these States to integrate the review process. Even in the 25 agencies with integrated reviews, the review process excludes GA, Social Security, and SSI, which may also be affected by reporting errors that directly affect food stamps. Thus, even integrating reviews cannot be viewed as sufficient to ensure that a net cost approach is implemented comprehensively.

- The increased complexity of reviews could adversely affect QC review completion rates. With scarce administrative resources, and the possible difficulties associated with determining the correct benefits in other programs, it is possible that the net federal cost approach could lead to a decline in the rates of review completions. Since higher error rates are attributed to incomplete reviews, this effect could paradoxically increase the State's official error rate.

It should be noted, finally, that a legislative change to certification policy could also avoid overstating the total cost of errors across multiple programs without complicating QC procedures. "Standardizing" income counted in food stamp certifications for AFDC and SSI households, demonstrated in Oklahoma as a way to simplify the certification process, would

make food stamp benefits less sensitive to changes in unreported income that also affect AFDC or SSI benefits.50/ Under such a policy, the AFDC or SSI maximum benefit for a given household size is counted as income in the food stamp certification calculations, regardless of the amount of reported unearned or earned income that affects the grant, since the sum of outside income and the AFDC or SSI benefit is constant.51/

If food stamp QC reviews were performed under this redesigned program, the calculation of the correct food stamp benefit would no longer include a separate consideration of the direct effects of earned or unearned income and the AFDC benefit (unless the amount of actual income would make the household totally ineligible for AFDC). Income-reporting errors would always be balanced exactly by the offsetting AFDC error. This type of legislative change would simplify food stamp certification for large segments of the caseload, and, in the process, would mitigate the potential overstatement of the total cost of errors for multiple-program households.

CONCLUSIONS

- Current procedures measure the loss to the Food Stamp Program given the actual income available to each household. The amount of the food stamp error is based on a calculation of the correct food stamp allotment using the actual amount of benefits issued and available to the household from other programs (AFDC, SSI, etc.) as countable income. An alternative approach would use the amount of AFDC or SSI income that would have been issued in the absence of the error. This alternative approach could be viewed as a measure of the net cost to the government of errors that affect food stamps and other programs. Using

---

50/The implementation of this policy variation is described in James Ohls et al., "Final Report for the Food Stamp Program Simplified Application Demonstration Evaluation," Mathematica Policy Research, Inc., September 1986.

51/Such a policy must, as in Oklahoma, provide for an adjusted income figure if the household is receiving the AFDC "30 + 1/3" earned-income deduction that is allowed from the first four months of an AFDC recipient's earnings.

the net cost approach, however, could introduce new food stamp errors where none existed, and increase administrative complexity and burden.

==========

Should Underissuances and Negative Errors Be in the Error Rate?

Current law focuses primarily on the dollar value of positive errors--overissuances and issuances to ineligible households. The QC process routinely measures and reports on underpayments to eligible households and improper negative actions, but fiscal liabilities are determined solely by the overpayment-error rate. This focus on overpayments could conceivably create an unbalanced set of incentives that encourage State agencies to emphasize avoiding overissuances and tolerating underissuances.

FNS reviewed the relationship between overpayments and underpayments in an effort to determine whether any evidence could be found that an emphasis on reducing overpayments contributes to an increase in underpayments.52/ This analysis used QC data for eight semiannual review periods from 1979 through 1983. The analysis suggested that the error-rate liability system had not had any demonstrable effect on underpayment error rates. Over- and underpayment error rates were generally found to move together over the period, suggesting that administrative performance affects both positive and negative errors similarly. On a national basis, no statistically significant negative correlation was found between overpayment and underpayment error rates. Only a few individual States showed correlations between the level of over- and underpayment rates, and these were mixed between positive and negative correlations.

Nevertheless, a review of the design of the QC system should consider feasible ways to ensure that the system provides appropriate incentives for avoiding of all types of certification errors. FNS examined the feasibility of incorporating underissuances and negative action errors in the official error rate.

52/Office of Analysis and Evaluation, Food and Nutrition Service, "The Relationship between Overpayment and Underpayment Error Rates in the Food Stamp Program: A Preliminary Analysis," November 1985.

Inclusion of Underissuances in the Error Rate. Underissuances are detected, measured, regressed, and reported as part of the same review process and are based on the same review sample as overpayment errors. Thus, a consistently defined error rate that incorporated overpayment errors and underissuances could be calculated on the basis of data already collected and current procedures.

Underissuances could be included in the official error rate by adopting a "total payment error rate"--the final payment-error rate plus the absolute value of the State's underissuance rate.53/ This measure holds States accountable for errors in both directions and eliminates any incentive that States may have to avoid overissuances by issuing food stamps too conservatively.

Adding a new component of error would increase the error rates of all States. In 1985, the national average underpayment-error-rate was 2.2 percent, ranging from a low of 0.5 percent in Nevada to a high of 4.2 percent in Oklahoma. The average underpayment-error rate has been fairly stable since the inception of QC, rarely rising above 2.5 percent. If such a change were contemplated, it might be appropriate to consider this expansion when error-rate liability thresholds are established. The effect of including underissuances in the error-rate measure would vary widely according to how these thresholds are set. Table III.6 shows the effect of adding underissuances to the error rate using several alternative assumptions about the level of the threshold. Column 2 shows the actual liability computed for Fiscal Year 1985 under current law, based on a 5 percent threshold, without the inclusion of underissuances in the payment-error rate. Columns 3, 4, 5, and 6 show the liabilities that would be computed with underissuances included in the payment-error rate, but with thresholds of 5, 6, 7, 8, and 10 percent. A 6 percent threshold would increase overall liabilities by 31 percent, and an 8 percent threshold would reduce overall liabilities by 31 percent.

Inclusion of Negative Action Errors in the Error Rate. Including negative action errors in the official error rate is more difficult than including underissuances. Negative action reviews are currently undertaken for a separate sample, drawn from a universe of specific actions (terminations and denials)

53/This approach was taken in the 1980 amendments to the Food Stamp Act and served as the basis of liabilities in 1981 and 1982.

Table of Contents

TABLE III.6

EFFECT OF INCLUDING UNDERISSUANCE IN ERROR RATE,
UNDER VARIOUS ADJUSTED ERROR RATE TARGETS:
ESTIMATED FY 1985 LIABILITIES

FISCAL YEAR 1985

State	Current Policy (5 percent threshold)	Including Underissuances with Liability Threshold of: 5 Percent	6 Percent	7 Percent	8 Percent	10 Percent
Northeast						
Connecticut	$ 1,025,885	$ 1,709,808	$ 1,025,885	$ 683,923	$ 341,962	0
Maine	598,696	1,396,957	997,827	598,696	399,131	0
Massachusetts	5,860,198	9,208,882	7,534,540	5,860,198	4,185,856	1,674,342
New Hampshire	0	136,631	43,876	0	0	0
New York	16,280,441	48,841,322	37,987,695	27,134,068	16,280,441	5,157,875
Rhode Island	391,265	912,952	652,109	391,265	260,844	0
Vermont	410,263	574,368	410,263	246,158	164,105	0
Mid-Atlantic						
Delaware	246,819	411,365	246,819	164,546	82,273	0
Washington, D.C.	1,561,937	2,900,739	2,454,472	2,008,204	1,561,937	669,401
Maryland	2,531,992	4,219,987	2,531,992	1,687,995	843,997	0
New Jersey	5,829,207	10,492,573	8,160,890	5,829,207	3,497,524	1,165,841
Pennsylvania	11,709,304	18,400,335	15,054,820	11,709,304	8,363,789	3,345,516
Virginia	1,415,766	4,955,179	3,539,414	2,123,648	1,415,766	0
Virgin Islands	299,390	470,470	384,930	299,390	213,850	85,540
West Virginia	111,525	509,248	254,624	0	0	0
Southeast						
Alabama	13,118,714	16,617,037	14,867,875	13,118,714	11,369,552	7,871,228
Florida	2,432,062	6,080,156	3,648,093	2,432,062	1,216,031	0
Georgia	16,441,248	29,088,361	26,558,938	24,029,516	21,500,095	16,441,248
Kentucky	776,939	2,330,816	1,553,877	776,939	0	0
Mississippi	1,816,892	5,450,675	4,239,414	3,028,153	1,816,892	605,631
North Carolina	1,802,557	4,506,394	2,703,836	1,802,557	901,279	0
South Carolina	8,319,451	12,159,197	10,879,282	9,599,366	8,319,451	5,759,620
Tennessee	2,058,553	3,087,829	2,058,553	1,029,276	0	0
Midwest						
Illinois	9,029,457	16,253,023	12,641,240	9,029,457	5,417,674	1,805,891
Indiana	5,659,493	8,174,823	6,917,158	5,659,493	4,401,528	1,886,498
Michigan	4,563,908	10,649,119	7,606,514	4,563,908	3,042,605	0
Minnesota	3,218,388	5,057,466	4,137,927	3,218,388	2,298,848	919,539
Ohio	3,690,595	6,150,991	3,690,595	2,460,396	1,230,198	0
Wisconsin	1,267,661	3,802,983	2,957,876	2,112,768	1,267,661	422,554
Southwest						
Arkansas	1,242,979	2,900,285	2,071,632	1,242,979	828,653	0
Louisiana	7,719,113	12,130,035	9,924,574	7,719,113	5,513,652	2,205,461
New Mexico	1,620,452	2,916,814	2,268,633	1,620,452	972,271	324,090
Oklahoma	5,312,273	10,034,294	8,853,789	7,673,284	6,492,779	4,131,768
Texas	28,120,597	40,618,640	34,369,619	28,120,597	21,871,576	9,373,532
Mountain-Plains						
Colorado	1,354,275	2,979,404	2,437,694	1,895,984	1,354,275	541,710
Iowa	2,028,618	2,840,065	2,028,618	1,217,171	811,447	0
Kansas	1,078,122	1,940,620	1,509,371	1,078,122	646,873	96,497
Missouri	487,902	1,610,024	805,012	0	0	0
Montana	385,539	899,592	642,566	385,539	257,026	0
Nebraska	1,152,601	1,481,915	1,152,601	823,286	493,972	164,657
North Dakota	0	0	0	0	0	0
South Dakota	0	0	0	0	0	0
Utah	583,204	972,006	583,204	388,802	194,401	0
Wyoming	138,332	345,830	207,498	138,332	69,166	0

TABLE III.6 (continued)

State	Current Policy (5 percent threshold)	Including Underissuances with Liability Threshold of: 5 Percent	6 Percent	7 Percent	8 Percent	10 Percent
West						
Alaska	2,096,708	2,733,341	2,487,537	2,241,733	1,995,929	1,504,321
Arizona	4,329,756	6,803,902	5,566,829	4,329,756	3,092,683	1,237,073
California	13,136,972	33,509,605	27,114,642	20,719,680	13,136,972	1,534,791
Guam	27,912	55,824	27,912	0	0	0
Hawaii	0	188,071	0	0	0	0
Idaho	57,098	230,521	53,530	0	0	0
Nevada	0	0	0	0	0	0
Oregon	3,800,149	5,971,663	4,885,906	3,800,149	2,714,392	1,085,756
Washington	4,048,211	6,361,475	5,204,843	4,048,211	2,891,580	1,156,632
U.S. Total	**$201,189,415**	**$372,073,609**	**$7,937,338**	**$9,040,785**	**$3,731,229**	**$71,167,012**

SOURCE: Ellen Kisker, "The Use of Alternative Measures of Error in Calculating FSP Liabilities," Mathematica Policy Research, Inc. (forthcoming).

rather than from the universe of active households. Moreover, negative action errors do not necessarily entail incorrect denials or the termination of benefits, because the negative action review is procedural; the QC reviewer does not determine the correct eligibility or benefit outcome.54/ In many cases, negative action errors are a failure by the eligibility worker to document the basis for denial or termination properly, or to allow the full prescribed 30 days for applicants to provide information before denial. In such cases, a negative action error does not necessarily mean that the household was in fact eligible and actually lost benefits.

Combining data on negative action errors with data on errors from the active household sample is thus problematic. Developing a combined error rate that incorporates negative action errors would entail creating an inevitably complex method to interpret very different measures. Moreover, treating negative action errors as issuance errors would overstate the true magnitude of applicants' loss of benefits, given the procedural nature of the review. Arriving at a more accurate measure of lost benefits would require broadening the scope of the negative action review to investigate actual eligibility and to compute benefits.

Current law does offer a precedent for treating the negative action error rate separately. States are now eligible for a higher federal match of their administrative costs if the combined overpayment and underpayment error rate is less than 5 percent and the negative action error rate is less than a separate threshold. Current procedures set this separate target equal to the weighted national average negative action error rate in the previous year. Some variant of this approach could be applied to calculating State liabilities as well.

CONCLUSIONS

- No evidence exists that underpayment-error rates have increased as States seek to reduce overpayment-error rates.

54/Food Stamp Program regulations require States to determine whether or not a loss of benefits results from an improper decision to deny or terminate food stamps. If so, the household is entitled to restored benefits. This process, however, is outside the scope of QC.

- It is feasible to include underissuances in the official error rate, using current samples of active households and data currently collected. They could be included by adding the absolute value of underissuance errors to the value of overpayment errors. Adding this new component of error would increase overall errors by 1 to 3 percentage points in most States.

- Negative action errors are measured using different samples and procedures than is true of payment errors, and focus on procedural errors. Negative action errors do not necessarily identify eligible households that were incorrectly denied benefits, since no attempt is made in the negative action review to determine actual eligibility or benefit amounts. Monitoring negative action errors can be accomplished through a separate review process.

D. ADJUSTMENTS TO STATE ERROR RATES

Under current policy, States are judged on the absolute levels of their errors. However, some have argued that States should be judged on the quality of their performance given the administrative difficulties they face. States are different in a number of obvious ways. Some have suggested that these differences have important implications for measuring errors. QC error rates may reflect not only the administrative performance of an agency, but also the difficulty of the caseload served and the characteristics of the operating environment--factors that, it could be argued, are beyond the control of State and local managers.55/ If some factors can be identified that have a clear and measurable effect on the difficulty of preventing errors, it might be appropriate to adjust error rates or, equivalently, error thresholds to account for the differences among States. Such adjustments would then, in theory, provide a better measure of relative State performance. Error rates would reflect the particularly difficult obstacles faced by some States, and would "reward" those that have succeeded in lowering error rates despite particularly difficult obstacles. This section examines the feasibility of identifying factors that have a systematic effect on error rates and adjusting error rates in an equitable and methodologically sound manner.

==========

Question D1: Do differences in the complexity of the caseload or local area characteristics explain some of the differences in State error rates?

Finding: Extensive analyses of the variation in errors identified a set of case characteristics and local area socioeconomic characteristics that make a statistically significant contribution to observed differences in the incidence or amount of errors. The characteristics, however, do not show up in all models as contributors to error. Case characteristics that affect the

---

55/This distinction is not always clear. Managers who face large, dynamic caseloads, for example, are likely to adopt a different set of administrative practices than are managers who face small, stable caseloads. While these managers cannot control the size or stability of the food stamp population, they can control their response.

probability or amount of errors are earnings, household size, the receipt of AFDC and SSI, the total number of medical, dependent-care, and shelter deductions, and the presence of certain assets. The only socioeconomic characteristic found to have a significant effect on the incidence of error was population density.

A major class of potentially important variables could not be included in this analysis. Measures of caseload dynamics--the rates at which households enter or leave the program and the frequency of changes in household circumstances--are simply not available. In their absence, only part of the effect of caseload characteristics on error is captured.

Question D2: Is there an empirical basis for adjusting State error rates for differences in caseload complexity or local area complexity?

Finding: The analysis does not provide a clear basis for adjusting error rates. In a purely statistical sense, many of the alternative adjustment models developed in this analysis performed equally well. The explanatory power was roughly the same, the correlation between different adjustments was reasonably high, and the regression coefficients were significant. Thus, there is no empirical way to select the "best" from among the range of alternatives.

Moreover, the various models tested produced error rate adjustments for particular States that varied, depending on the model used. The adjustments derived from the statistical models are quite sensitive to the choice of the included variables. For example, two models that are identical in all respects, with the exception that one includes a measure of population density and the other does not, produce adjustments that differ in direction for 9 States in 1984. Even for those States in which the direction of the adjustment is unchanged, the size of the adjustment changes. Because of the manner in which liabilities are determined under current law, even small changes in the adjusted error rates can have substantial fiscal consequences for some or all States.

The sensitivity of the adjustments to different specifications suggests that any selective focus on those factors that make intuitive sense may not improve equity. Choosing only those factors that are commonly associated with high error rates as the basis for an adjustment runs a sizable risk of introducing error into the QC measurement process. The adjusted rates in such a system would not necessarily be better, only different.

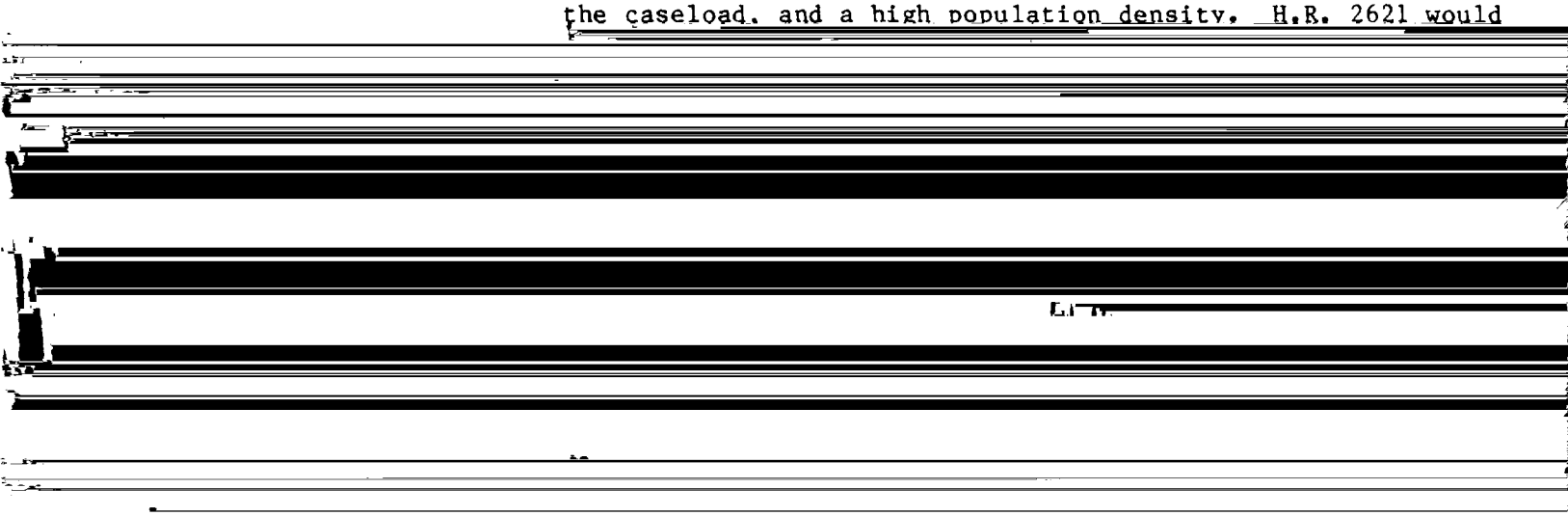
Moreover, the individual factors identified in this analysis affect error rates through a complex interaction. The effects of these factors frequently move in opposite directions and tend to offset each other. States vary according to a number of dimensions, some of which produce higher error rates and others lower. In most States, the factors that make the caseload more difficult to manage are balanced by other factors that make the caseload easier. Based on one model developed in this analysis, about three-quarters of the States would have had an adjusted 1984 error rate that was not significantly different from the actual, unadjusted error rate. Overall, the adjustment lowers the error rate for fewer than half the States and raises the error rate in the remaining States.

Should Adjustments Be Made for "Difficulty" or "Environment"?

If some States encounter more severe obstacles to accurate certification than others, they might be found to have higher QC error rates, even though their administrative resources and procedures are comparable to that found in States that exhibit lower measured error rates. If such were in fact the case, it could be argued that State error rates should be adjusted to reflect the effects of these factors so that the error rate provides a "better" measure of relative performance.

Approaches for adjusting error rates for caseload difficulty and the operating environment were formulated in two pieces of legislation introduced in the 99th Congress: H.R. 1279, introduced by Representative Matsui, and H.R. 2621, introduced

by Representative Jeffords.56/ H.R. 1279 would adjust AFDC liability thresholds upward based on three factors: the presence of an AFDC-U program, the incidence of earnings in the caseload, and a high population density. H.R. 2621 would



factors: the incidence of earned income in the caseload, the proportion of households with five or more members, the rate of caseload increase relative to the national average, and high or low population density.

The projected effects of these two bills demonstrate that the State-by-State impacts of error-rate adjustments are quite sensitive to the selection and definition of adjustment factors. Although these two bills have similar goals, and are similarly constrained in that they provide only for increases in thresholds (equivalent to reductions in error rates), they have quite different results. Table III.7 shows how these two bills would have changed Fiscal Year 1984 thresholds. On average, H.R. 1279 would increase the threshold by 0.4 percentage points; H.R. 2621 would increase the threshold by 0.7 percentage points. The proposals tend to favor different States; indeed, the correlation between the two sets of adjustments is only .32. Moreover, almost no correlation exists between these adjustments and the effects of the same factors that emerge from an analysis of QC data. This analysis developed statistical models of the actual effect of the factors identified in H.R. 1279 and H.R. 2621 on error. These results were then used to adjust State error rates in 1984. The correlation between these empirical adjustments and the legislative adjustments is extremely low: .03 and .08, respectively. These differences illustrate the difficulty of defining an equitable approach to adjusting error rates or thresholds based on "real world" operational experience, without considering the complex interrelationships among the many factors that can drive error rates up and down.

Does Caseload Complexity or Environment Make

To address this problem, a rigorous study was performed to determine whether statistically significant relationships exist between potential adjustment factors and error rates,

Table of Contents

TABLE III.7

ADJUSTMENTS TO 1984 STATE PAYMENT ERROR RATE THRESHOLDS
UNDER TWO LEGISLATIVE PROPOSALS

State	Actual FY 1984 Threshold ($)	Official Error Rate	H.R. 1279[a] Adjustment Due to . . . AFDC-U	Earnings	Population Density	Adjusted Threshold	Difference	H.R. 2621[b] Adjustment Due to . . . Earnings	Population Density	Household Size	Adjusted Threshold	Difference
Alabama	7.00	13.35	0.0	0.2	0.0	7.20	0.2	0.2	0.5	0.2	7.90	0.9
Alaska	10.45	9.29	0.0	0.1	0.0	10.55	0.1	0.1	0.5	0.1	11.15	0.7
Arizona	8.36	9.38	0.0	0.2	0.0	8.56	0.2	0.2	0.5	0.2	9.26	0.9
Arkansas	7.00	9.66	0.0	0.2	0.0	7.20	0.2	0.2	0.5	0.2	7.90	0.9
California	7.00	7.67	0.5	0.0	0.1	7.60	0.6	0.0	0.1	0.1	7.20	0.2
Colorado	7.25	10.66	0.5	0.0	0.1	7.85	0.6	0.0	0.1	0.1	7.45	0.2
Connecticut	8.04	7.11	0.5	0.0	0.0	8.54	0.5	0.0	0.1	0.2	8.34	0.3
Delaware	7.00	6.40	0.5	0.1	0.0	7.60	0.6	0.1	0.3	0.1	7.50	0.5
District of Columbia	7.93	8.80	0.0	0.0	0.5	8.43	0.5	0.0	0.5	0.2	8.63	0.7
Florida	7.48	8.95	0.0	0.1	0.0	7.58	0.1	0.1	0.3	0.1	7.98	0.5
Georgia	7.00	9.56	0.0	0.2	0.0	7.20	0.2	0.2	0.4	0.2	7.80	0.8
Guam	7.00	3.39	0.5	0.5	NA	--	--	0.5	NA	0.5	--	--
Hawaii	7.00	3.69	0.5	0.1	0.0	7.60	0.6	0.1	0.2	0.1	7.40	0.4
Idaho	7.00	6.88	0.0	0.2	0.0	7.20	0.2	0.2	0.5	0.2	7.90	0.9
Illinois[c]	7.00	8.31	--	--	--	--	--	--	--	--	--	--
Indiana	7.00	8.64	0.0	0.2	0.0	7.20	0.2	0.2	0.3	0.2	7.70	0.7
Iowa	7.00	8.51	0.5	0.2	0.0	7.70	0.7	0.2	0.5	0.2	7.90	0.9
Kansas	7.20	7.35	0.5	0.1	0.0	7.80	0.6	0.1	0.4	0.1	7.80	0.6
Kentucky	7.00	8.98	0.0	0.2	0.0	7.20	0.2	0.2	0.4	0.2	7.80	0.8
Louisiana	7.00	10.16	0.0	0.1	0.0	7.10	0.1	0.1	0.5	0.1	7.70	0.7
Maine	7.00	6.7	0.0	0.0	0.0	7.00	0.0	0.0	0.5	0.1	7.60	0.6
Maryland	7.91	6.85	0.5	0.0	0.5	8.91	1.0	0.0	0.5	0.2	8.61	0.7
Massachusetts	7.45	9.86	0.5	0.0	0.0	7.95	0.5	0.0	0.1	0.2	7.75	0.3
Michigan	7.00	6.46	0.5	0.0	0.3	7.80	0.8	0.0	0.3	0.3	7.60	0.6
Minnesota	7.00	9.77	0.5	0.1	0.0	7.60	0.6	0.1	0.2	0.1	7.40	0.4
Mississippi	7.00	9.24	0.0	0.2	0.0	7.20	0.2	0.2	0.5	0.2	7.90	0.9
Missouri	7.00	5.83	0.5	0.1	0.4	8.00	1.0	0.1	0.4	0.1	7.60	0.6
Montana	8.46	8.77	0.0	0.2	0.0	8.66	0.2	0.2	0.5	0.2	9.36	0.9
Nebraska	7.00	8.40	0.5	0.2	0.0	7.70	0.7	0.2	0.4	0.2	7.80	0.8
Nevada	7.00	2.54	0.0	0.0	0.0	7.00	0.0	0.0	0.5	0.1	7.60	0.6
New Hampshire	7.76	8.18	0.0	0.0	0.0	7.76	0.0	0.0	0.5	0.2	8.36	0.6
New Jersey	7.00	7.47	0.5	0.0	0.5	8.00	1.0	0.0	0.5	0.2	7.70	0.7
New Mexico	7.60	11.83	0.0	0.2	0.0	7.80	0.2	0.2	0.5	0.2	8.50	0.9
New York	8.34	10.14	0.5	0.0	0.0	8.85	0.5	0.0	0.2	0.3	8.84	0.5
North Carolina	7.00	7.22	0.0	0.2	0.0	7.20	0.2	0.2	0.5	0.2	7.90	0.9
North Dakota	7.00	6.27	0.0	0.2	0.0	7.20	0.2	0.2	0.5	0.2	7.90	0.9
Ohio	7.00	6.65	0.5	0.1	0.1	7.70	0.7	0.7	0.1	0.1	7.30	0.3
Oklahoma	7.00	7.61	0.0	0.1	0.0	7.10	0.1	0.1	0.5	0.1	7.70	0.7

Table of Contents

TABLE III.7 (continued)

State	Actual FY 1984 Threshold ($)	Official Error Rate	H.R. 1279[a] Adjustment Due to . . . AFDC-U	Earnings	Population Density	Adjusted Threshold	Difference	H.R. 2621[b] Adjustment Due to . . . Earnings	Population Density	Household Size	Adjusted Threshold	Difference
Oregon	7.00	9.18	0.0	0.3	0.0	7.30	0.3	0.3	0.5	0.3	8.10	1.1
Pennsylvania	7.00	10.41	0.5	0.0	0.5	8.00	1.0	0.0	0.5	0.1	7.60	0.6
Rhode Island	7.25	7.08	0.5	0.0	0.1	7.85	0.6	0.0	0.1	0.2	7.55	0.3
South Carolina	7.00	10.80	0.0	0.2	0.0	7.20	0.2	0.2	0.5	0.2	7.90	0.9
South Dakota	7.00	3.59	0.0	0.3	0.0	7.30	0.3	0.3	0.5	0.3	8.10	1.1
Tennessee	7.27	6.09	0.0	0.2	0.0	7.47	0.2	0.2	0.4	0.2	8.07	0.8
Texas	7.00	9.97	0.0	0.2	0.0	7.20	0.2	0.2	0.3	0.2	7.70	0.7
Utah	7.00	11.37	0.0	0.2	0.0	7.20	0.2	0.2	0.4	0.2	7.80	0.8
Vermont	7.00	9.71	0.5	0.1	0.0	7.60	0.6	0.1	0.5	0.1	7.70	0.7
Virgin Islands	8.32	7.63	0.0	0.7	NA	--	--	0.5	NA	0.5	--	--
Virginia	7.00	12.13	0.0	0.1	0.0	7.10	0.1	0.1	0.4	0.1	7.60	0.6
Washington	7.00	9.23	0.5	0.0	0.0	7.50	0.5	0.0	0.4	0.1	7.50	0.5
West Virginia	7.00	6.95	0.5	0.0	0.0	7.50	0.5	0.0	0.5	0.1	7.60	0.6
Wisconsin	7.04	9.60	0.5	0.1	0.2	7.84	0.8	0.1	0.2	0.1	7.44	0.4
Wyoming	7.17	9.08	0.0	0.3	0.0	7.47	0.3	0.3	0.5	0.3	8.27	1.1

[a] Adjustments are as follows: presence of AFDC-U, add 0.5 percent; for each 20% increment above national average of proportion of caseload with earnings, add 0.1 percent; and for each 20% increment above the national average on population density, add 0.1 percent.

[b] Adjustments are as follows: for each 20% increment above the national average of proportion of caseload with earnings and proportion of caseload with 5 or more members, add 0.1 percent; and for each 20% increment above or below national average on population density, add 0.1 percent. A final variable, rate of caseload increase, is not included because the food stamp caseload declined between 1983 and 1984. (H.R. 2621 defined the adjustment on the basis of growth relative to the national increases.)

[c] Because Illinois was involved in a demonstration during FY84, its error rates are not comparable to those of other States.

characteristics, and (3) individual State "indicator" variables to capture unmeasured administrative characteristics and State efforts to overcome difficult aspects of the caseload they serve or their environment.57/ The analysis attempted to identify characteristics associated with each review case that affected the likelihood of its having a QC error and the amount of the error, and thus the overall error rate for a State.

The analysis examined a wide range of local area characteristics, as listed in Table III.8. A total of 75 different case characteristics were also considered as potential factors that affect error rates:

- Household demographic characteristics, such as the age of household members, ethnicity, citizenship, and employment status
- Income, including earned and unearned income, and measured as incidence and amount
- Resources, including liquid and non-liquid assets
- Case actions, such as whether or not expedited service had been provided or work registration was required

A statistical model was constructed to estimate the effects of these three sets of variables (local area characteristics, caseload characteristics, and State indicators) on the payment-error rate. The payment-error rate is viewed in this model as comprising its component parts:

$$\text{Error rate} = \frac{(\text{Overissuance Rate} \times \text{Average Amount}) + (\text{Ineligible Rate} \times \text{Average Amount})}{\text{Total Allotments}}$$

The analysis entailed estimating the effects of various factors on each of the components of the overall error rate: the incidence of overissuances, the average amount of overissuances when they occur, the incidence of issuances to ineligibles, the average overissuance to ineligibles, and the

---

57/See Michael J. Puma and David C. Hoaglin, "The Effect of Caseload Characteristics and Socioeconomic Conditions on Food Stamp Payment Error Rates," Abt Associates, Inc. (forthcoming).

Table of Contents

TABLE III.8

GEOGRAPHIC AREA CHARACTERISTICS
USED IN AN ANALYSIS OF DIFFICULTY FACTORS

Variable Description
Whether household resides within one of 30 largest cities
Percent of housing units that are vacant
Percent of population that are Black
Percent of population that are Spanish
Percent of households that are families
Percent of households that are female-headed
Percent of persons 25 years of age or older with 12 or more years of education
Percent of civilian labor force that is unemployed
Percent of civilian labor force employed in agriculture
Percent of civilian labor force employed in manufacturing
Median family income
Percent of persons whose incomes are below poverty
Median gross rent for renter-occupied housing units
Total persons per square mile
Births per 1,000 resident population
Number of crimes per 100,000 resident population
Number of violent crimes per 100,00 resident population

allotment amount. The impacts of each component were then used as factors in an adjustment formula to derive adjusted error rates.

The characteristics with statistically significant effects in each of the separate models are summarized in Table III.9. It is clear that the characteristics were not uniformly important across the models. For the most part, six types of factors appear to affect food stamp errors:

1. Earned income tends to increase the incidence of food stamp errors and reduce food stamp allotments and error amounts.

2. Unearned income tends to increase the incidence of overpayment errors and reduce food stamp allotments and error amounts.

3. Household size tends to increase the incidence and amount of errors and to increase food stamp allotments.

4. Deductions tend to increase food stamp allotments and the incidence of overpayment errors.

5. Resources tend to increase the incidence of ineligibility errors.

6. Population density tends to increase the incidence of both overissuance and eligibility errors.

Is There a Basis for Adjusting Rates?

Given the findings reported above, the next question addressed by the analysis is whether an empirical basis exists for adjusting State error rates for factors with a statistical impact on error. This question is important because a complex adjustment process that is unstable or has only trivial effects on error rates would not improve the overall QC process as a basis for holding States accountable for their performance.

The results of an adjustment model for 1984 are shown in Table III.10. Error rates are both increased and decreased: the adjustments range from a reduction of nearly 3 percentage points in Pennsylvania to an increase of over 2 percentage points in California. The adjusted error rate is significantly different from the reported error rate in only 15 States.

Table of Contents

TABLE III.9

DIRECTION OF THE EFFECTS OF CASELOAD AND SOCIOECONOMIC CHARACTERISTICS ON FOOD STAMP ALLOTMENT, AND THE INCIDENCE AND AMOUNT OF PAYMENT ERROR

Variables	Allotment	Overpayment Error		Ineligibility Error	
		Incidence	Amount	Incidence	Amount
Household Size					
Number of Case Members	+		+		+
Number of Persons Ages 18-59		+		+	
Earned Income					
Reported Total Earned Income	-		-		-
"True" Presence of Earnings		+			
Reported Total Wage Income				+	
Unearned Income					
Total AFDC Grant	-		-		+
Total Unearned Income Other Than AFDC	-		-		-
Receipt of AFDC		+		-	
Occurrences of Institutional Unearned Income		+		+	
Occurrences of SSI				-	
Resources					
True Presence of Liquid Resources				+	
True Presence of Real Property				+	
True Presence of Vehicles				+	
Deductions					
Medical Deductions	+				
Dependent-Care Deductions	+				
Reported Shelter Costs	+				
Total Number of Deductions		+			
Population Density		+		+	

Table of Contents

TABLE III.10

COMPARISON OF 1984 STATE-REPORTED PAYMENT ERROR RATES TO ADJUSTED ERROR RATES

State	State-Reported Error Rates		Adjusted Error Rates		Difference
	Error Rate	Standard Error	Error Rate	Standard Error	(Adjusted-Reported)
Alabama	8.01	0.45	7.67	0.55	-0.34
Alaska	9.34	1.40	10.70	2.62	1.36
Arizona	9.58	0.49	10.99	0.73	1.41***
Arkansas	9.55	0.70	7.79	0.77	-1.76***
California	6.85	0.48	9.38	0.91	2.53***
Colorado	8.71	0.57	10.41	0.82	1.70***
Connecticut	7.01	0.62	8.36	0.96	1.35
Delaware	6.57	1.00	7.68	1.66	1.11
District of Columbia	8.91	0.91	6.27	1.38	-2.64***
Florida	7.69	0.44	8.12	0.62	0.43
Georgia	9.42	0.67	10.24	0.94	0.82
Hawaii	4.11	0.57	4.13	1.00	-0.02
Idaho	7.19	0.88	6.87	1.09	-0.32
Illinois	a	a	--	--	--
Indiana	8.51	0.57	8.41	0.71	-0.10
Iowa	8.22	0.66	8.30	0.80	0.08
Kansas	6.71	0.65	6.64	0.77	-0.07
Kentucky	8.91	0.54	9.02	0.71	0.11
Louisiana	10.05	0.72	9.90	1.05	-0.15
Maine	6.29	0.63	6.03	0.85	-0.26
Maryland	6.68	0.61	6.32	0.90	-0.36
Massachusetts	9.09	0.75	11.34	1.26	2.25**
Michigan	6.22	0.42	7.37	0.64	1.15**
Minnesota	8.61	0.66	8.63	0.83	0.02
Mississippi	8.19	0.64	7.97	0.74	-0.22
Missouri	6.20	0.37	5.41	0.43	-0.79***
Montana	8.43	0.83	8.95	1.16	0.52
Nebraska	8.64	0.70	7.51	0.76	-1.13***

TABLE III.10 (continued)

State	State-Reported Error Rates		Adjusted Error Rates		Difference
	Error Rate	Standard Error	Error Rate	Standard Error	(Adjusted-Reported)
Nevada	2.26	0.54	2.67	1.17	0.41
New Hampshire	8.31	1.00	8.64	1.31	0.33
New Jersey	7.26	0.46	7.75	0.61	0.49
New Mexico	10.25	0.58	10.33	0.72	0.08
New York	9.15	0.70	9.28	1.27	0.13
North Carolina	5.31	0.52	4.83	0.58	-0.48
North Dakota	6.31	1.10	3.87	1.11	-2.44***
Ohio	7.12	0.44	7.88	0.57	0.76**
Oklahoma	6.53	0.54	6.27	0.61	-0.27
Oregon	7.70	0.99	6.20	0.84	-1.50***
Pennsylvania	10.17	0.76	7.29	0.78	-2.88***
Rhode Island	7.23	0.60	7.43	0.79	0.02
South Carolina	7.97	0.53	9.44	0.70	1.47***
South Dakota	3.63	0.62	3.48	0.69	-0.15
Tennessee	5.82	0.53	5.71	0.71	-0.11
Texas	7.15	0.63	7.13	0.68	-0.02
Utah	9.64	1.03	8.82	1.42	-0.82
Vermont	8.99	0.92	8.07	0.95	-0.92***
Virginia	6.56	0.61	5.82	0.85	-0.74
Washington	9.18	0.49	9.75	0.68	0.57
West Virginia	6.08	0.57	6.58	0.75	0.50
Wisconsin	8.32	0.43	8.44	0.58	0.12
Wyoming	9.01	1.40	8.46	1.47	-0.55

[a] Illinois was involved in a demonstration during FY84. Thus, its error rates are not comparable to those of other States.

** Significantly different at the .05 percent significance level.
***Significantly different at the .01 percent significance level.

The manner in which these factors affect the components of the error rate and offset each other can be illustrated with the results for two States, California and Pennsylvania, as shown in Table III.11. This table shows the separate effects of the adjustment factors that have significant relationships to the five components of the model: the allotment amount, the incidence of overissuance, the amount of overissuance, the incidence of issuances to ineligibles, and the amount of issuances to ineligibles.58/

The data presented in Table III.11 show how the individual factors interact to produce a net adjustment. Take the adjustment to the average allotment in California as an example. If California had a caseload with characteristics identical to the national average, California's adjusted allotment amount per household would be about $3.00 higher than the reported value. The $3.00 is the net effect of the interaction among seven factors. For example, differences from the national average in terms of the incidence of earned and other unearned income and the average household size in California tend to reduce the adjusted allotment amount. This is more than offset by the upward adjustment due to California's high AFDC benefits (relative to the national average).

58/For the models that deal with the allotment amount and incidence of errors (overpayment or ineligible), the sum of the individual factor adjustments is approximately equal to the total adjustment (i.e., the difference between the reported and adjusted values). For the two models that focus on the amount of errors, however, the adjustments cannot be summed because the underlying equations are nonlinear. Thus, Table III.11 shows, on the "total" line, the total adjustment for each State based on the model, not the sum of the individual factor adjustments.

Table of Contents

TABLE III.11

EFFECTS OF ADJUSTMENTS FOR
ENVIRONMENT AND CASELOAD CHARACTERISTICS
ON ERROR RATE COMPONENTS FOR SELECTED STATES

	California	Pennsylvania
Allotment ($)		
Total Earned Income (-)	-7.9	-3.8
Medical Deductions (+)	+0.3	+0.02
Dependent Care Deductions (+)	+0.1	+0.08
Amount of AFDC Benefits (-)	+49.2	-0.6
Amount of Other Unearned Income (-)	-21.9	+5.4
Number of Persons in Case (+)	-14.7	+6.9
Shelter Costs (+)	-1.9	-0.4
Total	+$3.00	+$8.00
Incidence of Overpayment (%)		
Population Density (+)	-0.7	-2.0
True Presence of Income (+)	+0.3	+0.2
Number of Deductions (+)	+0.6	0
Occurrences of Institutional Unearned Income (+)	+0.8	+0.1
Receipt of AFDC (+)	-0.5	0
Number of Persons Ages 18-59 (+)	-0.3	-0.1
Total	+0.6%	-1.4%
Amount of Overpayment ($)		
Total Earned Income (-)	-0.02	-0.01
Amount of AFDC Benefits (-)	+0.20	-0.002
Amount of Other Unearned Income (-)	-0.06	+0.015
Number of Persons in Case (+)	-0.04	+0.017
Total[a]	+$5.40	-0.70
Incidence of Ineligibility (%)		
Population Density (+)	-0.4	-1.2
Amount of Wage Income (+)	+0.2	0
Receipt of AFDC (-)	+0.6	0
Occurrences of SSI (-)	-0.3	0
Occurrences of Institutional Unearned Income (+)	+0.4	0

TABLE III.11 (continued)

	California	Pennsylvania
Incidence of Ineligibility (%) (continued)		
Number of Persons Ages 18-59 (+)	0	0
True Presence of Liquid (+) Resources (+)	0	0
True Presence of Real Property (+)	+0.1	-1.6
True Presence of Vehicles (+)	0	0
Total	+0.8%	-2.4%
Amount of Ineligibility		
Total Earned Income (-)	-0.04	-0.02
Amount of AFDC Benefits (-)	+0.24	-0.003
Amount of Other Unearned Income (-)	-0.16	+0.04
Number of Persons in Case (+)	-0.08	+0.04
Totals[a]	+$25.80	+$12.40

[a]Individual adjustments are not additive because the form of the model is exponential.

In the same fashion, the reported incidence and amounts of errors are adjusted upward, but all of these changes are the net effects of both negative and positive adjustments due to individual factors.59/

The major point of this illustration is that adjustments to State error rates reflect a complex interaction among a number of factors commonly thought to be beyond the control of State and local managers. States vary according to a number of dimensions, some of which tend to increase error rates while others lower them. In most States, the factors that tend to increase error rates are balanced by other factors that tend to lower them. This means that a selective focus on a few factors might actually make comparisons among States less accurate.

The results presented thus far were based only on one of many adjustment models developed from the same data during the course of this analysis. In a purely statistical sense, many of these earlier variants performed equally well: the explanatory power was roughly the same, the correlation between the different adjustments was reasonably high, and the regression coefficients were significant. Thus, there is no empirical way to select the "best" from among the range of alternatives.

59/The role played by the level of AFDC benefits in explaining variation in food stamp error rates is illustrated by these results. Where AFDC payment standards are high and food stamp benefits correspondingly low, the effect is to lower the denominator in the formula for the food stamp payment-error rate, which yields a higher error rate; the reverse is true in States with low AFDC payment standards. In a special analysis of this interaction, it was shown that taking AFDC payment standards into account as an error-rate adjustment factor would increase error rates in 37 States and lower error rates in 15. (See Nancy Burstein, "Impact of AFDC Payments on Food Stamp Payment Error Rate," Abt Associates, Inc., forthcoming. The results were obtained for 52 States because Illinois was not included in this analysis. The District of Columbia, Guam, and the Virgin Islands were included.) For some States, as for California, exceptionally high or low AFDC benefits can generate an important adjustment to components of the payment-error rate. As with all of the other factors found to be significant individually, however, this factor may be largely offset by other factors that contribute opposing adjustments.

However, a sensitivity analysis of the modeling results reported above indicates that the adjustments are quite sensitive to the choice or specification of explanatory variables used, and to year-to-year differences in QC samples. For example, two adjustment models were estimated, identical in all respects with the exception that one included a measure of population density and the other did not. These two models produced adjustments that differed in direction for 9 States. Even for those States in which the direction of the adjustment was unchanged, the size of the adjustment changed. Population density appears to be an important variable in explaining differences among States, and its exclusion may well distort the final adjustments.

A strong possibility exists that the exclusion of other, equally important factors may have the same distorting effect. Of greatest concern is the absence of measures of caseload dynamics--the rates at which households enter and leave the program and the frequency with which their household circumstances change. Their inclusion would almost certainly result in different adjustments for some States. These measures simply do not exist and cannot be incorporated into an adjustment. In their absence, there is no way to know whether the adjusted error rates based on existing data are better than the unadjusted rates or simply different.

The potential for year-to-year variation in adjustment results also presents a problem. To be acceptable, an adjustment procedure should be relatively stable from one year to the next. Wide swings in the size or direction of error-rate adjustments would probably raise serious questions about the fairness or usefulness of the entire adjustment process. Thus, the models used in the error-rate analysis were estimated with both Fiscal Year 1984 and Fiscal Year 1985 QC sample data.60/ This analysis found ten States with large changes (i.e., more than one percentage point) in the adjustments from one year to the next. In seven of these ten States, the direction of the adjustment changed, up in one year and down in the other. Twenty States experienced moderate changes (more than half but less than a full percentage point). And in another 20 States the difference

---

60/Because population density figures for office areas could not be linked to the 1985 QC data, population density was omitted from the models in comparing 1984 and 1985 results. The basic conclusion about year-to-year variation, however, is not affected by this departure from the model results reported earlier.

between the 1984 and 1985 adjustment was less than half a percentage point. In some cases, these differences are large enough to make substantial financial differences. For example, New York's error rate would have been adjusted upward by 0.8 percentage points in 1984 but by 2.3 points based on the 1985 sample. The difference in the adjustment would have meant a difference of $5 million in financial liability.

Given the results of the analysis to date, which have been based on currently available data and modeling techniques, there appears to be no empirical foundation for adjusting error rates to account for the difficulty of program administration or the socioeconomic environment of the program. Although considerable progress has been made in identifying factors that contribute to variations in error rates, there is no empirical way to choose between competing specifications. Moreover, the results of the analysis are too sensitive to relatively minor changes in the definition of explanatory variables to provide an equitable and stable basis for adjusting error rates.

==========

CONCLUSIONS

- Extensive analyses of the variation in error rates identified a set of factors that have significant effects on the components of the error rate: the incidence of overissuances and issuances to ineligibles, the amount of overissuances and issuances to ineligibles, and the average allotment amount. These factors include the presence of earnings, the receipt of AFDC and other types of unearned income, household size, deductions, the presence of certain assets, and population density. This implies that there may be a potential to develop a better measure of the relative performance of States. Building such an adjustment model, however, presents serious problems. A major class of potentially important factors--measures of caseload dynamics--could not be included because of data limitations. In their absence, the effect of the measured variables must be interpreted with some caution.

- Analyses indicated that errors are affected by a complex interaction of variables that offset or enhance one another to varying degrees across and within States. States vary in a number of ways, some of which lead to higher error rates, while others lead to lower rates. This means that adjusting error rates by a few indicators with prespecified

adjustments may not improve the measure of a State's error rate.

- The analysis to date does not provide a clear basis for adjusting State error rates. Many adjustment models perform equally well, and there is no empirical way to select the "best" from among them. Moreover, the error-rate adjustments are very sensitive to the choice of the variables included. Because of the manner in which liabilities are determined under current law, even small changes in the adjusted error rates can have substantial fiscal consequences.

==========

IV. POLICY ISSUES: ERROR RATE TARGETS AND FINANCIAL LIABILITIES

Under the current law, an important application of error-rate information is to provide a basis for judging State performance relative to a legislated standard, and for establishing State liabilities for failure to meet the defined performance standard. Defining a performance standard and establishing a method for imposing financial liabilities are inherently policy issues. In contrast, the issues discussed in Chapter III, although they interact with these policy issues, pertain largely to technical issues about methods: how best to define error and collect data on its occurrence, so that the resulting data depict the performance of States as accurately and fairly as possible.

However, issues associated with performance standards and liabilities pertain to goals and incentives, and will ultimately be resolved by policy debate rather than technical analysis. This chapter focuses on three broad questions about standards and liabilities:

****

MAJOR POLICY ISSUES

Issue A: How, and at what level, should the performance standard or threshold used to establish liabilities be set?

Issue B: What method should be used to compare State error rates with the performance standard in order to decide whether or not a State has met the standard?

Issue C: How should the amount of liability be calculated?

****

In this chapter, the broad issues concerning error-rate targets and fiscal liabilities are divided into three groups. Just as with the more technical issues discussed earlier, any amendment of current QC policy must recognize the interactions among the issues and the potential tradeoffs. Choices about each aspect of the system can enhance or offset the consequences of other design decisions. For example, the threshold could be set at a point lower than current law, but with no change to total liabilities if the rate at which liabilities are assessed is altered to account for the lower threshold. Alternative methods for defining a liability threshold level are discussed in Section A. In Section B,

alternative approaches for making decisions on assessing liabilities are presented, and, in Section C, various ways to calculate the amount of the liability are examined. A consideration of possible changes to the QC system, however, must include a careful examination of the joint effects of these particular issues.

A. ESTABLISHING A PERFORMANCE STANDARD

Congress has established a target for the food stamp payment-error rate--5 percent--which all States must meet or risk facing financial liabilities for excessive errors. This policy establishes an "arbitrary" or "judgmental" threshold, in the sense that no empirical analysis has established 5 percent as the "right" target. The 5 percent target reflects a political consensus. It accepts the notion of some unavoidable level of error in food stamp certifications, but still sets incentives to make substantial improvement over the current performance of most States. The current 5 percent target is the final step in a phased-in threshold reduction sequence mandated by the 1980 and 1982 amendments to the Food Stamp Act.

Although error rates have declined in recent years, they have not fallen as rapidly as error-rate targets; thus, performance-based liabilities have increased dramatically. This increase in liabilities has given rise to renewed debate over liability thresholds and the appropriate methods for comparing performance with targets and thus identifying States subject to liabilities. This section of Chapter IV examines two questions about liability thresholds:

==========

Question A1: What methods could be used to establish liability thresholds, and what are their advantages and disadvantages?

[redacted]

Finding: No sound empirical method exists for setting the level of liability thresholds. Setting thresholds based on the cost-effectiveness of further error reduction is not practical because it is infeasible to predict or measure the benefit savings due to various error-reduction steps

ity into the error-rate target. Setting error-rate targets is most appropriately a policy decision that rests on tradeoffs among (1) accurate program administration, (2) the recognition of some level of unavoidable error, and (3) the appropriate allocation of the cost of issuance errors between States and the federal government.

Question A2: What is the appropriate relationship between food stamp liability thresholds and thresholds in other benefit programs?

Finding: Comparisons between the characteristics of the food stamp and AFDC caseloads can be considered in any evaluation of the relative thresholds in these two programs. Similarly, differences in the complexity of program rules and program funding incentives can be taken into account.

Differences in caseload characteristics explain almost all of the difference between food stamp and AFDC error rates. Program-specific rules account for a large portion of errors in both programs.

No compelling empirical evidence exists to suggest that the food stamp liability threshold should match the threshold used in the AFDC program.

==========

How, and at What Level, Should Liability Thresholds Be Set?

Five basic approaches could be taken in setting thresholds. Each has its own potential advantages and disadvantages, administrative implications, and associated incentives for program performance.

A zero threshold approach is one in which any error rate greater than zero leads to a liability. A zero threshold would give States strong incentives to avoid issuing any erroneous benefits. It implies financial liabilities even for those States that exhibit very low error rates, and implies a decision to extend financial incentives to reduce errors to all States.

A relative threshold approach sets a threshold based on the overall pattern of the recent error-rate performance of all States. For example, the threshold for liabilities could be set each year at the level of the lowest observed error rate in that year or the previous year, in which case all States but one would be subject to liabilities. Alternatively, the threshold could be set at some higher point in the distribution of error rates (e.g., the 25th percentile, or the median error rate). Setting the threshold on such a basis would be supported by evidence that the target is attainable, but would most likely always leave at least some States subject to performance liabilities.61/ This method also has the disadvantage of creating error-rate targets that are unstable. Goals would be adjusted each year, and a particular State might be subject to a liability in one year and not face a liability in another year, even with the same error rate in both years.

A State-specific threshold sets an individual threshold for each State based on its operating environment. A single national threshold confronts all States with the same standard, regardless of possible differences across States in terms of the difficulty, cost, or potential impact on other program goals of reducing their error rates to the standard. In the past, State agencies have raised questions about whether the particular operating circumstances of each State should be considered in setting its error-rate target.
This point of view is a corollary of arguments for adjusting error rates to reflect differences across States in terms of caseload characteristics or program environment. Adjusting State liability thresholds based on the severity of such factors is an alternative to adjusting each State's observed error rate before comparing it with a standard threshold, as described in Chapter III. As pointed out in Chapter III, however, a far-ranging analysis that succeeded in identifying factors that affect error rates does not provide a clear basis for selecting an adjustment model. Equally good, alternative models produce sharply different adjustments for the error rates of different States. The same problems confront any attempt to adjust error-rate targets.

61/In theory, for example, all States could meet a target set at the 25th percentile in the distribution of the previous year's error rates, but it is unlikely that all States with higher error rates the previous year could improve their performance sufficiently in one year to avoid liabilities the next year.

A "benefit-cost" approach, for the nation as a whole or for individual States, would set targets at the point at which the cost of reducing error rates is judged to exceed the savings from reducing issuance errors further. State targets would be adjusted to ensure that they encourage error reduction only when it is cost-effective to do so (i.e., when the savings that will accrue from a given reduction in error are expected to exceed the administrative cost of achieving the error reduction).

There is no reasonable prospect that a sound method can ever be developed to adjust State thresholds based on the cost-effectiveness of error reduction. Although demonstrations have been conducted to test the effects of some (but not all) error-reduction strategies, the design of these demonstrations precluded any clear conclusions about cost-effectiveness, or any generalization of the results to other States or other strategies. The cost-effectiveness of particular corrective action strategies is likely to vary according to each State's error rate, the methods it has already applied to prevent errors, and the local administrative factors that affect the cost of error-reduction efforts. Moreover, new strategies are being introduced continuously, and any assessment of cost effectiveness would thus quickly become out of date. It is very unlikely, therefore, that it will be possible to develop a standard approach that would equitably adjust error-rate targets for all States based on a consistent analysis of each State's options for reducing errors.

An arbitrary or judgmental threshold approach sets targets at a level judged by policymakers to be politically acceptable. The present 5 percent threshold is based on an "arbitrary" rather than an empirical process.

There is no available empirical evidence that can be used to develop the "right" threshold, nor any evidence that the results of the alternative methods described above would be more acceptable to State or federal agencies or more equitable to the relative interests of different States. The level at which a national threshold should be set remains a policy question about how the cost of program errors should be allocated fairly between State and federal governments.

==========

CONCLUSIONS

- No theoretically or empirically superior method exists for establishing liability thresholds. Analysis can suggest the relative merits of different threshold levels, but cannot develop an unequivocal, correct level.

==========

How Should the Food Stamp Threshold Compare with the AFDC Threshold

Although no "correct" liability threshold can be derived from available data and analytical methods, information is available which can provide a context for setting the liability threshold in the Food Stamp Program both in terms of overall complexity of the program and relative to the threshold used in the AFDC program. Food stamp error rates have been consistently higher than AFDC error rates. In Fiscal Year 1984, for example, these rates for the nation as a whole were 8.6 percent and 6.0 percent, respectively--a difference of 2.6 percentage points. Several analyses were conducted to determine the reasons for this difference.

The first set of analyses illustrates the contribution that complex aspects of Food Stamp Program design make to current error rates.62/ These analyses, based on Fiscal Year 1984 QC data, estimated the effect of "defining away" errors by using two approaches for simplifying food stamp eligibility policy. The first was a simulation of the effect on error rates of instituting a "standard benefit policy" for "pure" AFDC/food stamp households, under which food stamp benefits would be determined from a simple table based only on household size, the presence of earnings (but not the amount), and the presence of an elderly or disabled member.63/ The second

---

62/The analyses described below are contained in James Ohls and Jennifer Schore, "Potential Effects of Program Changes on Food Stamp Program Error Rates," Mathematica Policy Research, Inc. (forthcoming).

63/"Pure" AFDC/food stamp households are those in which all individuals participate in both programs. A policy of this sort was in fact implemented and evaluated in the Illinois Simplified Application Demonstration. A similar simplifying policy was implemented by Oklahoma in the same demonstration. Those results are discussed in Chapter III of this report.

study examined the potential effect of eliminating the shelter cost deduction on error rates.

As shown in Tables IV.1 and IV.2, such design simplification could reduce current food stamp error rates, although the advantages of simplification would have to be weighed against the "benefit-targeting" purposes of current rules, such as the shelter cost deduction. For example, applying the "standard benefit policy" nationally would have resulted in a 12 percent decline in overpayment-error rates (from 8.6 to 7.6 percent). Eliminating the shelter deduction would have reduced the payment-error rate by almost 7 percent (from 8.6 to 8.1 percent).

Another way of thinking about where the threshold should be set is to compare the AFDC and Food Stamp Programs. Are the problems that face food stamp administrators different from those that face other programs? If they are, it would provide some indication about where the threshold should be set. Although both food stamp and AFDC error rates have declined, error rates remain consistently higher in the Food Stamp Program, as shown in Table IV.3 for FY 1984. Reported payment-error rates were significantly higher (at the 10 percent level) for food stamps than AFDC in 40 of the 51 State agencies for which the comparison could be made. In no instance was the AFDC reported payment-error rate significantly higher than the food stamp rate.

State agencies, however, also face different thresholds for financial liabilities in the two programs: a 5 percent payment-error rate threshold in food stamps and 3 percent in AFDC. To review relative thresholds and error rates in the two programs, it is essential to identify as clearly as possible the sources of error-rate differences. Three hypotheses have been proposed to account for differences in error rates between the two programs:

1. Differences in client characteristics. The population served by the FSP may be more difficult to administer correctly, particularly because it contains more cases with earnings.

2. Differences in program design complexity. The Food Stamp Program design may be more complex, and therefore more difficult to administer without error.

3. Differences in program administration. States may devote more care to administering the AFDC program than the FSP, both because they have a greater sense of "owning" their AFDC programs and because they share the cost of the AFDC benefits with the federal government.

Table of Contents

TABLE IV.1

ESTIMATED 1984 ERROR RATE REDUCTION
UNDER A STANDARD BENEFITS POLICY

	Household Type		
	Pure AFDC	Other	All Households
Payment Error[a]			
1984 Rate	5.3[b]	11.0[b]	8.6
Estimated Percent Reduction	48.2%	-	12.2%
Adjusted 1984 Rate	2.8	11.0	7.6
Underpayment Error			
1984 Rate	1.7	2.7	2.3
Estimated Percent Reduction	50.4%	-	15.3%
Adjusted 1984 Rate	0.8	2.7	2.0
Case Error			
1984 Rate	19.5	25.1	23.4
Estimated Percent Reduction	55.0%	-	14.3%
Adjusted 1984 Rate	8.8	25.1	20.0

NOTE: Analysis is based on a weighted sample of case data from the July-August 1984 quality control sample.

[a]Includes overissuances and issuances to ineligibles.

[b]Weighted rates are derived from national average of State-regressed QC error rates.

Table of Contents

TABLE IV.2

ESTIMATED 1984 ERROR RATE REDUCTION
WITH THE ELIMINATION OF SHELTER DEDUCTION

	Household Type		
	Pure AFDC	Other	All Households
Payment Error[a]			
1984 Rate	5.3[b]	11.0[b]	8.6
Estimated Percent Reduction	9.7%	5.7%	6.8%
Adjusted 1984 Rate	4.8	10.4	8.1
Underpayment Error			
1984 Rate	1.7	2.7	2.3
Estimated Percent Reduction	31.0%	10.2%	16.5%
Adjusted 1984 Rate	1.2	2.5	2.0
Case Error			
1984 Rate	19.5	25.1	23.4
Estimated Percent Reduction	28.9%	13.8%	17.7%
Adjusted 1984 Rate	13.9	21.7	19.2

NOTE: Analysis is based on a weighted sample of case data from the July-August 1984 quality control sample.

[a] Includes overissuances and issuances to ineligibles.

[b] Weighted rates are derived from national average of State-regressed QC error rates.

Table of Contents

TABLE IV.3

STATE-REPORTED PAYMENT ERROR RATES FOR
FOOD STAMPS AND AFDC - FY 1984

State	Food Stamps	AFDC
Alabama	8.01	2.73
Alaska	9.34	9.28
Arizona	9.58	11.30
Arkansas	9.55	3.90
California	6.85	3.27
Colorado	8.71	3.53
Connecticut	7.01	3.13
Delaware	6.57	5.44
District of Columbia	8.91	10.54
Florida	7.69	1.07
Georgia	9.42	6.54
Guam	4.22	-
Hawaii	4.11	5.38
Idaho	7.19	5.56
Illinois[a]		
Indiana	8.51	4.02
Iowa	8.22	3.48
Kansas	6.71	4.22
Kentucky	8.91	3.98
Louisiana	10.05	5.87
Maine	6.29	4.14
Maryland	6.68	5.40
Massachusetts	9.09	6.95
Michigan	6.22	8.92
Minnesota	8.61	2.07
Mississippi	8.19	2.50
Missouri	6.20	3.63
Montana	8.43	4.81
Nebraska	8.64	5.86

Table of Contents

TABLE IV.3 (continued)

State	Food Stamps	AFDC
Nevada	2.26	1.78
New Hampshire	8.31	6.37
New Jersey	7.26	3.78
New Mexico	10.25	6.30
New York	9.15	5.60
North Carolina	5.31	3.01
North Dakota	6.31	2.03
Ohio	7.12	5.95
Oklahoma	6.53	2.80
Oregon	7.70	3.68
Pennsylvania	10.17	10.11
Rhode Island	7.23	3.40
South Carolina	7.97	7.45
South Dakota	3.63	2.54
Tennessee	5.82	4.09
Texas	7.15	4.85
Utah	9.64	6.36
Vermont	8.99	4.28
Virgin Islands	11.64	1.90
Virginia	6.56	2.38
Washington	9.18	3.48
West Virginia	6.08	4.19
Wisconsin	8.32	5.76
Wyoming	9.01	3.89
TOTAL	7.83	5.13

[a]Illinois was involved in a food stamp demonstration during FY84; hence, its error rates are not comparable to other States.

A second set of studies was conducted to address these issues. The analyses and results for each hypothesis are described below.64/

Differences in Client Characteristics. Testing the hypothesis about the relative difficulty of administering food stamp and AFDC caseloads required answering the following question: What would the food stamp payment error rate be if the food stamp caseload were identical to the AFDC caseload?

The results of this simulation analysis indicate that case characteristics indeed account for virtually all of the difference between the payment-error rates of the two programs. As shown in Table IV.4, in 29 of the 47 States where the 1984 reported food stamp payment-error rate exceeded the AFDC rate, the simulation pulls the food stamp error rate below the actual AFDC error rate. In other words, the observed difference between food stamp and AFDC error rates is fully explained in those 29 States by differences in caseload characteristics.65/ In 17 of the remaining States in which the food stamp error rate was higher, the simulation yielded a lower food stamp payment-error rate, but it remained above the actual AFDC error rate.66/ For these States, case characteristics explain part, but not all, of the present excess of food stamp error rates above AFDC rates.

For the nation as a whole, case characteristics account for all of the difference between food stamp and AFDC payment error rates. This simulation indicates that, if the Food Stamp Program were serving the same population as the AFDC program, its error rate in FY 1984 would have been about 4.0 percent, compared with the AFDC rate of 5.1 percent.

Two differences between the food stamp and AFDC caseloads were identified as major contributors to the error-rate dif

---

64/This analysis is reported in Nancy Burstein, Marie Hojnacki, and Kaye Husbands, "Differences Between Food Stamp and AFDC Error Rates," Abt Associates, Inc. (forthcoming).

65/In an additional 4 States, actual AFDC error rates exceeded actual food stamp rates, and this difference is accentuated further when food stamp error rates are simulated for the AFDC population.

66/In one State, Alaska, the simulation yielded a higher food stamp error rate.

Table of Contents

TABLE IV.4

SIMULATED FOOD STAMP PAYMENT ERROR RATES (PER)
FOR AFDC RECIPIENTS

State	Actual Reported 1984 Food Stamp PER (1)	State Reported 1984 AFDC PER (2)	Simulated Food Stamp PER for AFDC Cases (3)	Percent of AFDC/Food Stamp Difference Explained by Case Characteristics (4)
Alabama	8.01	2.73	3.21	91
Alaska	9.34	9.28	10.25	0
Arizona	9.58	11.30	3.77	100
Arkansas	9.55	3.90	5.30	75
California	6.85	3.27	4.38	69
Colorado	8.71	3.53	5.30	66
Connecticut	7.01	3.13	3.54	90
Delaware	6.57	5.44	2.62	100
District of Columbia	8.91	10.54	6.04	100
Florida	7.69	1.07	2.17	83
Georgia	9.42	6.54	4.80	100
Hawaii	4.11	5.38	2.96	100
Idaho	7.19	5.56	3.91	100
Indiana	8.51	4.02	1.58	100
Iowa	8.22	3.48	4.50	79
Kansas	6.71	4.22	3.82	100
Kentucky	8.91	3.98	4.26	94
Louisiana	10.05	5.87	3.35	100
Maine	6.29	4.14	1.90	100
Maryland	6.68	5.40	5.29	100
Massachusetts	9.09	6.95	7.50	74
Michigan	6.22	8.92	4.97	100
Minnesota	8.61	2.07	5.29	51
Mississippi	8.19	2.50	3.47	83
Missouri	6.20	3.64	2.87	100
Montana	8.43	4.81	8.12	8
Nebraska	8.64	5.86	5.06	100
Nevada	2.26	1.78	0.57	100
New Hampshire	8.31	6.37	6.01	100
New Jersey	7.26	3.78	2.85	100
New Mexico	10.25	6.30	5.49	100
New York	9.15	5.60	3.95	100
North Carolina	5.31	3.01	1.83	100
North Dakota	6.31	2.03	4.46	43
Ohio	7.12	5.95	4.47	100

Table of Contents

TABLE IV.4 (continued)

State	Actual Reported 1984 Food Stamp PER (1)	State Reported 1984 AFDC PER (2)	Simulated Food Stamp PER for AFDC Cases (3)	Percent of AFDC/Food Stamp Difference Explained by Case Characteristics (4)
Oklahoma	6.53	2.80	2.18	100
Oregon	7.70	3.68	2.89	100
Pennsylvania	10.17	10.10	3.55	100
Rhode Island	7.23	3.40	2.20	100
South Carolina	7.97	7.45	4.41	100
South Dakota	3.63	2.54	1.55	100
Tennessee	5.82	4.09	2.43	100
Texas	7.15	4.85	4.36	100
Utah	9.64	6.36	7.22	74
Vermont	8.99	4.28	4.96	86
Virginia	6.56	2.38	0.73	100
Washington	9.18	3.48	5.60	63
West Virginia	6.08	4.19	2.10	100
Wisconsin	8.32	5.76	5.80	99
Wyoming	9.01	3.89	2.96	100
Virgin Islands	11.64	1.90	1.47	100
TOTAL	7.83	5.13	3.96	100

NOTE: Column (4) is calculated as ((1) - (3)) / ((1) - (2)). It is bounded by 0 and 100 percent. This value is shown as 100 percent for States in which the simulated FSP PER is less than the AFDC PER as well as for States where the calculated ratio would be greater than 100 percent.

ferences. First, the Food Stamp Program serves additional types of households (single individuals, two-parent families, childless couples, and elderly persons) that exhibit higher error rates than the remainder of the food stamp caseload; these additional household types explained almost half of the difference between the actual food stamp error rate and the simulated food stamp error rate for the AFDC caseload (about 1.6 percentage points). Second, even for food stamp households that are generally similar to AFDC cases (approximately one-half of all households), differences in composition--most notably the proportion of households with earned income--contribute to higher food stamp error rates. This difference increases the food stamp error rate by 2.2 percentage points.

Program Design Complexity. Additional analysis was conducted to examine the possibility that differences in program complexity also contribute to differences in error rates between AFDC and food stamps. If one program's legislative and regulatory requirements are more complex, eligibility staff might find it more difficult to understand and apply them, which might contribute to higher observed error rates.

This analysis focused on the frequency with which errors occurred (1) in both programs, (2) only in AFDC, and (3) only in food stamps. The analysis was based on Fiscal Year 1984 QC error findings for cases with identical case composition that received both food stamps and AFDC from the 29 States for which integrated review results were available. Focusing the analysis on this population ensured that case characteristics did not contribute to differences in food stamp and AFDC error findings, since the samples examined for the two programs were identical. The analysis assumed that, because these "pure" AFDC/food stamp cases are handled by a single eligibility worker or benefit from the information collected for AFDC, differences in program administration had no effect.67/ However, because of the analytical focus and sample, the quantitative findings from the analysis must be interpreted with caution.

The results indicate that program-specific regulations are a large factor in QC errors. As shown in Table IV.5, 18 percent

---

67/There are in fact some agencies in which separate workers handle eligibility determination for the two programs, even for such pure AFDC/food stamp cases. To some extent, therefore, differences in program administration could have a marginal effect on these findings, attributable to differences in the complexity of policy.

of the 11,000 cases had errors, but only a third of those (6 percent) had errors in both programs. Moreover, some cases with errors in both programs actually had multiple variances that reflected different errors in the two programs. Consequently, only 19 percent of all variances that were found actually applied to both programs; 81 percent of all variances arose from program rules peculiar to one or the other program.

Cases with food stamp errors only accounted for 7 percent of all the sample cases, and AFDC-only errors for 5 percent, as shown in Table IV.5. An analysis of the differences between AFDC and food stamp regulations, and of the incidence of particular error types, helped identify the regulations peculiar to each program that contribute to the surprisingly large incidence of single-program errors. In the Food Stamp Program, program-specific rules pertaining to the excess shelter deduction and unearned income accounted for most of the errors that affect only food stamps. In AFDC, program-specific rules pertaining to basic program requirements (deprivation, child support, and work registration) and needs and case requirements contribute to the majority of AFDC-only errors.

Program-specific regulations governing both programs thus contribute heavily to the incidence of QC errors. This analysis suggests that the effect of program-specific rules may be slightly more pronounced in the Food Stamp Program, because the incidence of program-specific errors was somewhat higher than the comparable effect in AFDC. Thus, the design complexity of the Food Stamp Program may be seen as contributing marginally to the observed difference between food stamp and AFDC error rates, although this contribution is not as pronounced or certain as the effects attributable to differences in caseload characteristics.

Administrative Effects. A third factor that could help explain higher food stamp error rates is the different administrative structure of AFDC and food stamps. A hypothesis is sometimes advanced that the States' share in the cost of AFDC benefits creates a stronger incentive to reduce or avoid errors than in the Food Stamp Program, where benefits are entirely federally financed. Responsibility for a share of AFDC benefit payments, according to this hypothesis, can lead State agencies to devote more careful attention or more administrative resources to the AFDC eligibility process.

Available data are insufficient to allow a conclusive analysis of administrative effects. However, exploratory analyses provide at least some indirect suggestion of possible effects. Error findings were analyzed for two sets of food

Table of Contents

TABLE IV.5

CASE ERROR FINDINGS BY PROGRAM

Food Stamp Error Status	AFDC Error Status: No Overpayment	AFDC Error Status: Overpaid/Ineligible
No Overpayment	82% (9,114)	5% (558)
Overpaid or Ineligible	7% (755)	6% (658)

NOTE: Cases examined are "pure" AFDC/food stamp households from States with integrated QC reviews in Fiscal Year 1984.

stamp cases drawn from the 1984 QC data: (1) "pure" AFDC/food stamp cases (those that receive benefits from both programs, and in which the case composition is identical for the two programs); and (2) "lookalike" cases (those that receive only food stamps, but fit the same case composition criteria).68/ Cases that receive AFDC as well as food stamps could be expected to experience the benefits of AFDC administration, in that any information discovered by the AFDC program would probably be made available to the food stamp program as well.69/

Since the "lookalike" cases did not receive AFDC, however, they were obviously different from those that did. As a group, they had more income, primarily earned, from sources other than AFDC.

The results of the analysis do not suggest that participation in AFDC leads to a case's receiving greater administrative attention and thus being less prone to error. When a multivariate analysis controlled for household size and the receipt of earned and unearned income, it was found that the receipt of AFDC significantly increased rather than decreased the payment-error rate. The conclusion of this exploratory analysis is that, if differences in administration contribute at all to differences in AFDC and food stamp error rates, their effect is more likely to reduce food stamp errors relative to the AFDC rate.

68/For both the AFDC/food stamp cases and the "lookalikes," cases consisted only of persons who were categorically eligible for AFDC (i.e., a head of household and children under 17 in States with only AFDC-R programs, plus cases including a spouse of the head of household in States with AFDC-U programs).

69/In some instances, as pointed out earlier, pure AFDC/food stamp cases may not be administered by a single worker or experience the benefits of AFDC administration. However, if the effect of joint administration is in fact to lower error rates, than the separate administration of some AFDC/food stamp cases would have the effect of reducing the observed difference in error rates between the two groups of cases. The comparison between pure AFDC/food stamp cases and lookalikes can therefore be viewed as yielding a conservative estimate of the effect of administration.

CONCLUSIONS

- The overall analysis indicates that differences in AFDC and food stamp error rates are explained by differences in caseload characteristics and possibly to some extent by differences in the complexity of program design. Indirect analysis provides no evidence that food stamp error rates are higher because States devote more administrative resources to AFDC. Consequently, no compelling empirically based reasons exist for setting food stamp liability thresholds at the same level as thresholds for AFDC.

B. COMPARING MEASURED ERROR RATES WITH THE LIABILITY THRESHOLD

As pointed out in Chapter III, because the current design of the food stamp QC system provides essentially unbiased estimates of error rates, repeated samples yield an average error-rate estimate that is extremely close to the true error rate. It was also noted, however, that present sample sizes exhibit considerable sampling variability, so that the error rates estimated from particular samples can be expected to vary around the true value. Even with an unbiased procedure for estimating error rates, the level of precision attainable with current sample sizes leaves the possibility that comparisons between estimated State error rates and liability thresholds will yield different decisions than ones based on the true error rate in two areas: (1) whether or not a liability should be imposed, and (2) the amount of the liability.

Sampling variability means that it is possible for States to face liabilities based on any single sample that are quite different from the liabilities that would be calculated if the true error rate were known. For example, a State with a true error rate of 8 percent would have a 10 percent chance of incurring a liability as low as 10 percent or lower or as high as 35 percent or higher of the federal share of administrative costs. A State with a true error rate of 4 percent has about a 16 percent chance of being assessed a liability of 5 percent or more.70/

Given some unavoidable level of sampling error in the estimated State error rates, some risk always exists that the estimated liabilities will be different from the liabilities that would be determined if the true error rate were known. Allocating the cost of this risk is of critical importance. It is a delicate balance to strike. State agencies have

[redacted]

the consequences of sampling error, and adopt a method that

This section examines the general issue of how error-rate estimates should be used to judge State performance in a way which allocates the risks associated with sampling error appropriately between the federal and State governments. It examines the current method for comparing error-rate performance with the liability thresholds as a basis for assessing and calculating liabilities, as well as five alternative strategies.

==========

Question B1: How does the current procedure allocate the cost of sampling errors between federal and State agencies?

Finding: Using the point estimate will sometimes lead to a decision to assess or not to assess a liability against a State in any one year that would differ from an assessment decision based on the true error rate. This risk is due to sampling error. Under current law, this risk is somewhat greater for the States, because of a small, average overstatement of liabilities at all levels introduced by the step function. For States whose true error rates are near the threshold, this risk is greater. In this case, the error in the amount of the liability calculated under current law will be higher on average than for States whose error rates are higher. This is because error-rate estimates above the threshold lead to positive liabilities, but estimates below the threshold lead to no liability rather than to a negative value. For the current pattern of error rates--in which most States are substantially above the threshold--the potential bias is small for most States.

==========

How Does Current Policy Allocate the Cost of Sampling Errors?

A State's estimated error rate must be compared with the threshold in order to establish whether a State should be assessed a liability and to determine how large that liability should be. This is an area in which the variability or precision of the estimated error rate is very important. The imprecision inherent in the sampling process works in two directions; an approximately equal chance exists that a State's estimated error rate lies below or above the true

error rate. Thus, there is risk on both sides of the equation. A State faces a risk that it could be assessed a liability when it should not be or that its liability is too high. The federal government also faces a risk that States which should be assessed liabilities are not assessed or that the liabilities are too low. Since one of the criteria of a performance measurement system is fairness, it is important that these risks are not systematically borne on one side of the equation.

The probability of an erroneous decision to assess or not to assess a liability may not be trivial, depending on the relationship between a State's true error rate and the liability threshold, as well as on the magnitude of the sampling error.71/ This issue is important because current procedures lack symmetry. When the estimated error rate is above the threshold, a liability if assessed. However, when the estimated payment error rate is below the threshold, a counter balancing negative liability of credit is not given. The risk of such a liability decision increases as a State's true error rate approaches the threshold and as sampling error increases (as the sample size decreases).

This increased risk occurs because sample-based estimates are distributed around the true error rate for repeatedly drawn samples. Thus, roughly half the estimates from repeated sampling will be above the true error rate and roughly half below. In the case of a State whose true error rate equals 5 percent--and whose liability would be zero if the true rate were known--about half of the sample-based estimates will be above 5 percent and half below. Under current procedures, the State will be liable for a portion of the cost of excessive errors every time the estimated error rate is above the target but no offsets are given when the estimated rate is below the target. In this case, current procedures increase the cost of sampling error to the States.

This illustration assumes that the true error rate could be determined. In reality, of course, it can never be known without reviewing every case in a State's caseload. The issue, then, is whether the "risk" associated with these decisions is reasonable. For a State whose true error rate is high (say, 7 percent or more), as is currently the case for many of the States, the probability of an erroneous decision

---

71/The same can be said of the chance that the liability amount will be different. This issue is discussed further in Section IV.C.

to hold the State liable is very low. Table IV.6 shows for each State the relationship between the 1985 error-rate estimate and the lower bound of the two-tailed 90 percent confidence interval. There is only at most a 5 percent probability that the true error rate falls below this lower bound. In most cases, the lower confidence bound is above 5 percent. Thus, for most States, the probability that the true error rate was below the threshold in 1985 was very low.

What Other Methods Can Be Used to Make Liability Decisions?

Chapter III noted that the "acceptable" degree of sampling error is a policy choice that must consider the cost of obtaining greater precision and the consequences of sampling variability. This choice can be informed by examining a variety of strategies for comparing estimates from samples with a fixed standard where two conditions exist: the risk associated with sampling errors is judged to be too large, and the costs of increasing sample sizes are judged to be too great. For the purpose of this discussion, five strategies are illustrated in the context of the food stamp quality control system.

Accumulate Sample Results Over Time. A cumulative multi-year error rate could be computed each year. For example, the average of all error-rate point estimates for the past five years might be determined. Similarly, results might be accumulated over a longer or shorter period until the aggregated results reach a desired level of reliability. In the meantime, each year's liability could be based on that year's cumulative error rate, using the current procedure based on administrative costs.72/ A cumulative liability could then be computed with its own increasingly smaller sampling error and narrower confidence bound. Sampling error would be taken into account by recording a cumulative cash transfer for each year based on the lower bound of the confidence interval for the cumulative liability. In any given year, an actual cash transfer would be made to the federal government only if the cumulative cash transfer had risen since the previous year.

While this procedure is statistically sound and would substantially eliminate the problem of annual sampling error, accumulating sample results over time would complicate the liability computation and accounting process, increase administrative complexity, and make the determination of each year's liability difficult to understand. Moreover, the

---

72/Alternatives could also be developed in which the liability computation is based on benefits issued.

Table of Contents

TABLE IV.6

LIABILITIES UNDER ALTERNATIVE COMPUTATION METHODS

State	Payment Error Rate	Lower Bound of 90% Confidence Interval	Liability Under Current Policy (Point Estimate)	Liabilities Under Alternative Policies -------Based on Administrative Cost---------- Lower Bound	Point Estimate if Lower Bound Greater than .05
Alabama	0.1350	0.1194	$ 13,118,714	$ 9,620,390	$ 13,118,714
Alaska	0.1353	0.0830	2,096,708	810,883	2,096,708
Arizona	0.0938	0.0816	4,329,756	3,092,683	4,329,756
Arkansas	0.0788	0.0607	1,242,979	828,653	1,242,979
California	0.0708	0.0588	13,136,972	4,378,991	13,136,972
Colorado	0.0848	0.0692	1,354,275	541,710	1,354,275
Connecticut	0.0704	0.0567	1,025,885	341,962	1,025,885
Delaware	0.0717	0.0495	246,819	0	0
District of Columbia	0.0981	0.0797	1,561,937	669,401	1,561,937
Florida	0.0671	0.0554	2,432,062	1,216,031	2,432,062
Georgia	0.1291	0.1107	16,441,248	13,911,825	16,441,248
Guam	0.0533	0.0308	27,912	0	0
Hawaii	0.0435	0.0302	0	0	0
Idaho	0.0516	0.0294	57,098	0	0
Illinois	0.0816	0.0702	9,029,457	5,417,674	9,029,457
Indiana	0.1090	0.0957	5,659,493	4,401,828	5,659,493
Iowa	0.0841	0.0700	2,028,618	811,447	2,028,618
Kansas	0.0816	0.0610	1,078,122	431,249	1,078,122
Kentucky	0.0600	0.0516	776,939	534,763	776,939
Louisiana	0.0976	0.0825	7,719,113	5,513,652	7,719,113
Maine	0.0791	0.0638	598,696	399,131	598,696
Maryland	0.0737	0.0666	2,531,992	1,687,995	2,531,992
Massachusetts	0.0971	0.0803	5,860,198	4,185,856	5,860,198
Michigan	0.0735	0.0607	4,563,908	3,042,605	4,563,908
Minnesota	0.0951	0.0714	3,218,388	1,379,309	3,218,388
Mississippi	0.0798	0.0637	1,816,892	1,211,261	1,816,892
Missouri	0.0523	0.0454	487,902	0	0
Montana	0.0744	0.0639	385,539	257,026	385,539
Nebraska	0.0904	0.0744	1,152,601	493,972	1,152,601

Table of Contents

TABLE IV.6 (continued)

State	Payment Error Rate	Lower Bound of 90% Confidence Interval	Liability Under Current Policy (Point Estimate)	Liabilities Under Alternative Policies -------Based on Administrative Cost---------- Lower Bound	Point Estimate if Lower Bound Greater than .05
Nevada	0.0248	0.0174	0	0	0
New Hampshire	0.0442	0.0296	0	0	0
New Jersey	0.0850	0.0748	5,829,207	3,497,524	5,829,207
New Mexico	0.0883	0.0781	1,620,452	972,271	1,620,452
New York	0.0711	0.0576	16,280,441	5,426,814	16,280,441
North Carolina	0.0649	0.0503	1,802,557	61,500	1,802,557
North Dakota	0.0353	0.0221	0	0	0
Ohio	0.0743	0.0649	3,690,595	2,460,396	3,690,595
Oklahoma	0.1058	0.0846	5,312,273	2,951,263	5,312,273
Oregon	0.0941	0.0740	3,800,149	1,628,635	3,800,149
Pennsylvania	0.0936	0.0740	11,709,304	5,018,273	11,709,304
Rhode Island	0.0800	0.0660	391,265	260,844	391,265
South Carolina	0.1210	0.1062	8,319,451	5,759,620	8,319,451
South Dakota	0.0315	0.0193	0	0	0
Tennessee	0.0639	0.0527	2,058,553	761,209	2,058,553
Texas	0.1038	0.0872	28,120,597	15,622,554	28,120,597
Utah	0.0726	0.0532	583,204	126,604	583,204
Vermont	0.0806	0.0600	410,263	164,105	410,263
Virgin Islands	0.0973	0.0800	299,390	213,850	299,390
Virginia	0.0667	0.0570	1,415,766	707,883	1,415,766
Washington	0.0950	0.0830	4,048,211	2,891,580	4,048,211
West Virginia	0.0507	0.0423	111,525	0	0
Wisconsin	0.0800	0.0663	1,267,661	845,107	1,267,661
Wyoming	0.0678	0.0438	138,332	0	0
Total	0.0830		$ 1,189,415	$ 4,550,326	$ 200,119,826

impact of any one year's performance on fiscal liabilities would become known only through a complex process, which would tend to weaken the link between performance and fiscal consequences.

Adjust Liability Amount. Liabilities could be computed as under present policy, but also adjusted or "scaled down" to reflect the possibility that the true error rate is actually below the threshold. This downward adjustment of the liability amount would be greater for States whose estimated error rate is closer to the threshold. Adjusting liability amounts has some intuitive appeal, but departs from rigorous statistical methodology. Given the importance of maintaining a sound technical design for the quality control system and its use, this strategy is not assessed further.73/

Use the Lower Bound of the Confidence Interval. State agencies have argued that the federal government should make liability decisions by comparing the lower bound of a confidence interval with the threshold, rather than using the point estimate itself. If a sufficiently high confidence level is specified, this approach would virtually eliminate any risk that liabilities would be imposed due to sampling error when the true error rate is below the target, and would even eliminate liabilities for States with true error rates moderately above the target. Liabilities would be affected in two ways. First, States whose point estimates were above the threshold but whose lower confidence bound fell below the threshold would have their liabilities eliminated. All other States with liabilities would have them reduced. Table IV.6, displayed earlier, illustrates the effect of using the lower bound of a 90 percent (two-tailed) confidence interval--which implies that there is only a 5 percent chance that the true error rate is below the lower bound. Six States would have had their 1985 liabilities eliminated, and, overall, liabilities would have been reduced by 44 percent, from just over $200 million to about $110 million.

73/This alternative would entail adjusting liability amounts based on the probability that the true error rate is below the threshold. This probability could be estimated only by assuming that the distribution of error-rate estimates is known--that they are normally distributed around the point estimate. However, the assumption that the point estimate is in fact the true error rate around which estimates are distributed is invalid.

Substituting the lower bound for the point estimate, however, presents a distorted picture of error rates and true liabilities. Treating the lower confidence bound as the formal estimate of error rates would systematically and substantially understate true error rates. Using the lower bound as the basis for liability decisions would make total liabilities systematically lower than the total that would be imposed if true error rates were known. The federal government would absorb the difference. This approach would not improve equity; it simply shifts more of the risk to the federal government.

Moreover, using the lower bound could be viewed as potentially creating a perverse incentive for less accurate and fewer State reviews. As pointed out in Chapter III, discrepancies between State and federal case review results tend to increase the standard error of the official error rate, thus widening the confidence interval and lowering its lower bound. Smaller samples have the same effect. Wider confidence intervals, in effect, reduce performance standards, distort the QC system's representation of the incidence of true errors, and can in the extreme weaken incentives to reduce errors as a result.

Use the "Conditional Point Estimate." A less far-reaching change to current policy would continue use of the point estimate, except when the lower bound of the confidence interval falls below the liability threshold. The effect of such a policy would be to eliminate liabilities whenever a reasonable chance exists that the true error rate is below the threshold. States whose estimated error rates are above the threshold would avoid liabilities if the lower bound of their confidence interval fell below 5 percent. Table IV.6, presented earlier, illustrates the effect of such a policy on 1985 liabilities, assuming the use of a 90 percent two-tailed confidence interval.

The liabilities of the same six States would have been eliminated as in the previous example--those States with point estimates above 5 percent but lower bounds below 5 percent. However, in contrast to the universal use of the lower bound, the liabilities of other States whose point estimates are above the threshold are not affected. For 1985, overall liabilities are reduced only marginally (less than 1 percent) from the total determined under current policy.

Many of the same shortcomings of the lower bound apply here as well, although fewer States are likely to be affected. The factors that affect sampling error might be altered by the State, possibly leading to smaller samples, less precise results, and less accurate reviews. This strategy might also

weaken the incentives for real error reductions. As more States reduce their error rates and approach the targets, this strategy becomes the same as using the lower bound. An advantage of this alternative, from the standpoint of the States, is that they would be better protected against any substantial risk of the sampling error's affecting the basic decision to assess or not to assess a liability.

Use the Point Estimate with Credit for Low Error Rates. This strategy allows error rates below the applicable threshold to create "negative liabilities," or credits, that could be applied against future positive liabilities. This approach could make the liability determination process more equitable in two ways. First, it would allow exceptionally good performance to be recognized along with poorer performance. Second, it would mitigate the effects of sampling error for States whose error rates cross the liability threshold from year to year. Instead of treating the difference between the threshold and estimated error rates below it as zero, negative differences could be permitted. This would ensure that both the estimated error rate and the estimated liabilities are unbiased. Under current law, the estimated error rate is unbiased (as shown in Chapter III), but the estimated liability is biased upwards.

==========

CONCLUSIONS

- Although point estimates of State error rates are unbiased, the estimates may lead to an erroneous decision to assess or not to assess a financial liability for any single State in a given year. Given the currently high levels of errors relative to the threshold, the probabilities of this occurring are relatively small in most States. As State error rates approach the threshold, however, the current procedure for assessing liabilities creates an upward bias in estimates of the "true liability"--that which would be assessed if the true error rate were known. This upward bias would occur because current law does not permit a "negative" liability when, due to sampling error, the estimated error rate is less than the threshold in any given year.

- The point estimate reflects the best available measure of a State's performance, and is, therefore, the best measure on which to base liabilities. A lower confidence bound systematically and substantially understates the true liability (i.e., the liability that would be assessed if the true overpayment-error rate were known).

==========

C. METHOD FOR COMPUTING LIABILITY AMOUNTS

Once the error-rate estimate is found to exceed the threshold by a specific margin, the next issue is how that "excess error" should be converted into a liability. The liability computation method is an important choice for several reasons. First, the manner in which the liability relates to the error rate creates a message for both State agencies and the public at large about the purpose of the QC system. Second, it establishes particular types of incentives for reducing errors. Third, the liability calculation method defines how the cost of certification error is shared between the federal and State governments. Finally, the calculation method can affect the degree to which calculated liabilities constitute biased or unbiased estimates of the true liability--the liability that would be computed based on the true error rate. This section examines two central questions concerning the liability calculation method.

==========

Question C1: What effect does the present "step function" for calculating liabilities--with liabilities increasing in large increments at each percentage point boundary above the threshold--have on the fairness of the liability system?

Finding: The present step function for calculating liabilities produces potentially large differences in liability proportions for States which exhibit very similar error rates and administrative costs. Current procedures also overstate liability amounts--slightly for all States on average and more substantially for States whose true error rates are near the threshold. The result is that the liability is an average of 0.6 percent higher than the true liability for States whose true error rates are between 8 and 9 percent, and 2 percent higher for States whose true error rates are between 7 and 8 percent. A continuous function would link the liability to the actual estimated error rate rather than to the interval in which it falls. When used in conjunction with a credit system and the point estimate, this approach would distribute the risk of sampling errors equally between States and the federal government.

Question C2: What is the difference between calculating liabilities based on administrative cost, as under current policy, and calculating them based on an estimate of benefits erroneously issued?

Finding: Using an administrative cost base can penalize States that have devoted greater resources to administration, and could discourage greater expenditures on reducing errors.

==========

What Is the Effect of the Step Method for Calculating Liability?

Under present legislation, liabilities for States whose payment-error rates exceed the threshold are calculated on the basis of a "step function," as illustrated in Figure IV.1. Liabilities are set at a particular percentage of federal administrative reimbursement for all error rates that fall in specified one-percent intervals above the threshold. The liability "steps up" to a new liability proportion if the error rate crosses the boundary of an interval. For example, for all error rates from 5.01 to 6.00 percent, the liability is 5 percent of the federal administrative reimbursement; if the error rate is above 6.00 percent, the liability steps up to 10 percent.74/ The next one-percent interval carries a liability of 15 percent, and above 8 percent each interval increases the liability by an additional 10 percent (to 25 percent, 35 percent, and so forth). Using this step function raises issues about the equitable treatment of States that exhibit very similar circumstances, and about the unbiased nature of the liabilities that are computed.

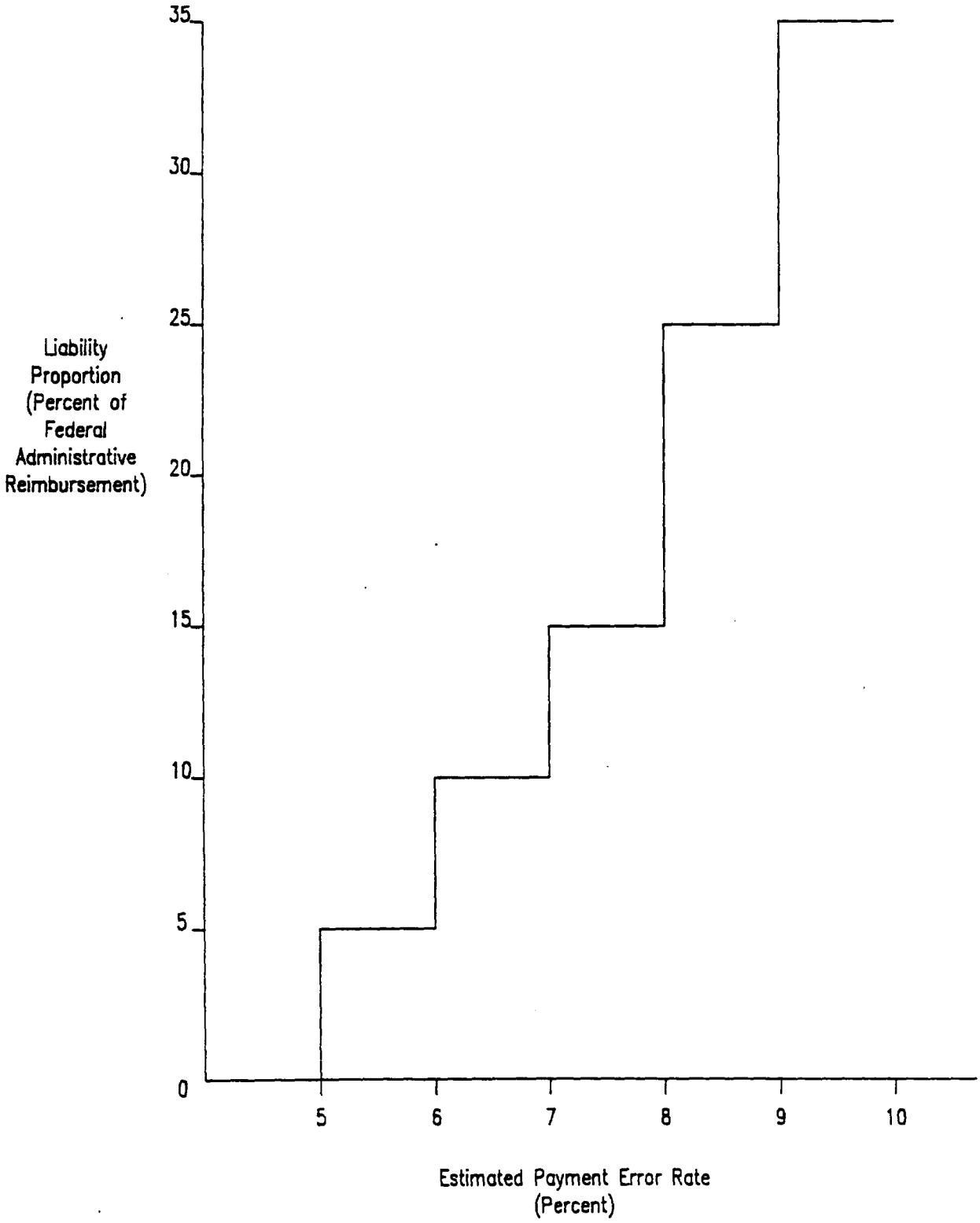
Given present methods for computing the amount of the financial liability imposed on a State, even trivial differences in error rates can cause large differences in liabilities. For example, a State whose error rate is 6.00 percent would be liable for 5 percent of the federal share of their administrative costs, but if its error rate were one-one-hundredth of a point higher (6.01 percent) its liability would be doubled, to 10 percent of administrative reimbursement. For States whose estimated error rate is close to one of the interval bound

---

74/The step function is equivalent to rounding all estimated error rates up to the nearest two-decimal error rate (e.g., 6.34 percent to 7.00, 7.86 to 8.00, etc.) and then basing liabilities on the difference between the rounded error rate and the threshold.

FIGURE IV.1

CURRENT METHOD FOR COMPUTING FINANCIAL LIABILITIES: THE "STEP FUNCTION"



aries, or "notches," in the step function, a review decision on even one case can thus have a substantial financial effect.

Even assuming that no sampling error exists and that the QC system could determine true error rates with perfect accuracy, States with very similar performance can face liabilities at quite different rates.

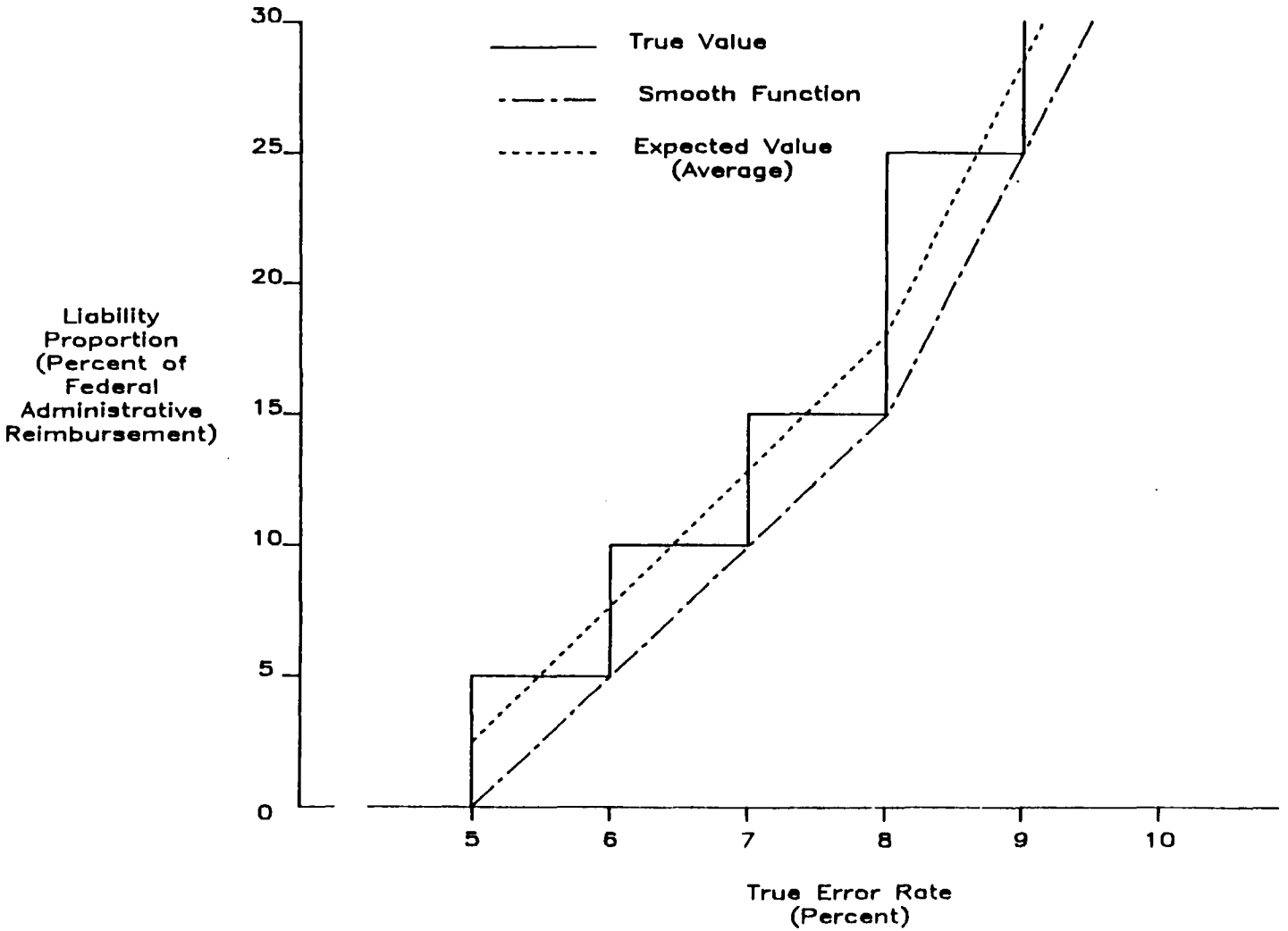
It is therefore open to question whether QC liabilities are related closely enough to variations in error rates. The issue here does not concern the ability of the QC system to measure errors accurately, but the volatility of the relationship between estimated error rates and the resulting liabilities. Rather than a smooth, continuous relationship in which higher error rates above the threshold are consistently associated with higher financial consequences, the current "step" method leads to large increases in liabilities for some States that exhibit very small differences in error rates, and significant ranges in which differences in error rates do not affect the liability amount at all. The step method thus creates different liabilities for States which exhibit very similar performance and similar administrative costs.

One desired statistical property for calculating liability under current law is that States whose true error rates are in the same interval should have their liabilities calculated the same way (i.e., their administrative reimbursement should be reduced by the same percentage). For example, if State A had a true error rate of 6.3 percent and State B had a true error rate of 6.6 percent, they should each have their administrative cost reimbursement reduced by 10 percent since they fall in the same interval, 6.0 to 6.99, in the step function (see Figure IV.1).

Much the same approach used in Chapter 3 to prove that the error measure itself is unbiased can be used to test for any potential bias in the liability calculation itself. Here, the test simulates the actual liability that would be calculated from repeated samples for those two States given true error rates of 6.3 and 6.6 percent. For the liability calculation to be statistically unbiased, the average liability calculated should be 10 percent of administrative cost reimbursement for both States. As can be seen in Figure IV.2, however, the average or expected value of the liability is not the same for States whose true error rates are above and below the midpoint of each interval. The average long-run liabilities are indicated by the diagonal line running close to but somewhat above the midpoint of each step. This means that State A, with a true 6.3 percent error rate, is under-assessed in the long-run at a reduction of less than 10 percent. Conversely,

Table of Contents

FIGURE IV.2
COMPARISON OF THE CURRENT LIABILITY CALCULATION METHOD AND A SMOOTH FUNCTION



State B, with a true error rate of 6.6 percent, is assessed above the 10 percent rate on average.

Thus, liability amounts are systematically overstated for some States and understated for others, depending on where the true error rate of a State falls within one of the step function intervals. For true error rates below the midpoint of a step interval, the procedure yields, on average, an underestimate of the true liability proportion. For true error rates above the midpoint of a step interval, the procedure yields an overestimate of the true liability proportion on average. Moreover, on average, there is a slight upward bias. In any specific case, of course, there is no way to know whether an upward or downward bias exists, because the true error rate is not known.

A smooth liability formula would avoid both problems--"notches," or large differences in liability associated with small error-rate differences, and bias in the estimation of true liability. If coupled with liabilities and "credits" of the sort described earlier, both based on the point estimate, the calculated liability would, on average, nearly equal the liability that would be assessed if the true error rate were known. In other words, this approach would distribute the risk of sampling error equally among States and the federal government. It could be calculated as follows:

$$\begin{aligned} \text{Liability Rate} &= 5(r - .05) && \text{if } r \text{ is between .05 and .08} \\ &= .15 + 10(r - .08) && \text{if } r \text{ is greater than .08} \\ &= 0 && \text{if } r \text{ is less than .05} \end{aligned}$$

This approach retains the basic structure of the current formula, but uses an unrounded estimate of the error rate instead of rounding up to the nearest percentage point. As shown in Figure IV.2, this formula would yield liability proportions that, on average, are lower than under current design.75/

The principle of a smooth, rather than a step, function for calculating liabilities would have the same advantages whether liabilities continue to be based on administrative cost or are based alternatively on benefits issued, the final issue discussed below.

75/By adding a constant, the formula could be revised to make it yield liabilities comparable to the expected value of the current step function.

CONCLUSIONS

- The current step function for determining financial liabilities presents two complex problems. First, the approach produces a systematic difference in the amount of liabilities computed relative to the liabilities that would be computed based on true error rates. The effect is to overstate liabilities in some instances and understate liabilities in others, but, on the average, to overstate them slightly. Second, the liability calculation introduces wide disparities in liability amounts for States with only trivially different error rates and ignores significant differences in error rates.

- A smooth liability formula avoids notches and bias in the estimation of true liability. A smooth function, when used in conjunction with a credit system and the point estimate, would distribute the risk of sampling errors equally between State and the federal government.

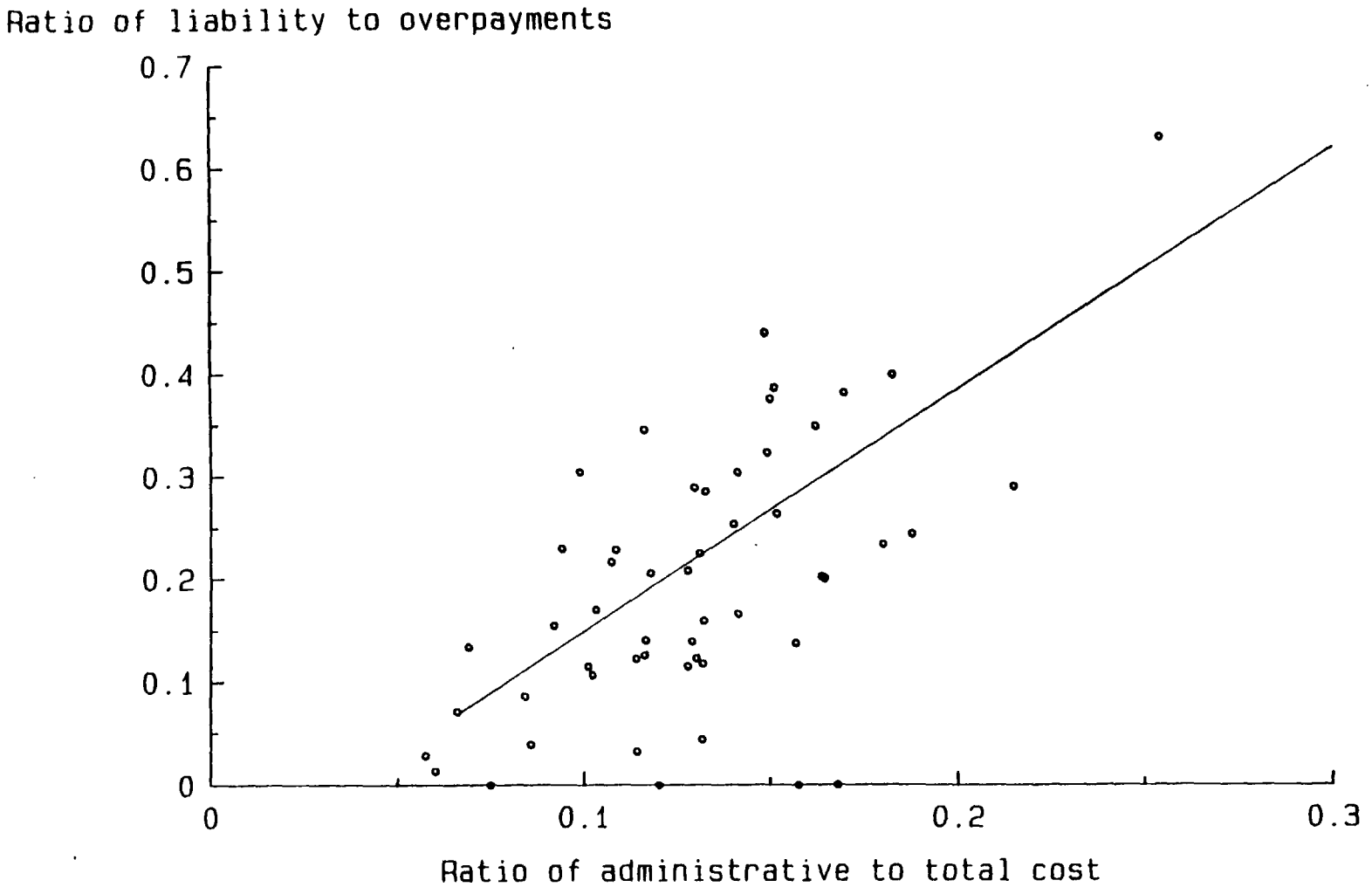
What Is the Difference between Basing Liabilities on Administrative Costs or on Erroneous Benefit Issuance?

The amount of financial liability for a State whose error rate is above the threshold depends on the State agency's administrative cost, although the amount of the liability may not exceed the dollar value of erroneous overissuances. Thus, if two States have identical error rates, caseloads, and benefit costs, the State that is spending more on administration will be assessed a higher liability. An issue has been raised about whether this method creates undesirable incentives to limit investments in reducing errors, or whether it unfairly affects States with higher administrative costs.

The differential effect of the current design is illustrated in Figure IV.3. The ratio of liability to total overissuance is plotted against the ratio of total administrative cost to total program cost (including administration and benefits).

Table of Contents

FIGURE IV.3
RELATIONSHIP OF LIABILITY TO ADMINISTRATIVE COST AND COST OF OVERPAYMENTS



This relationship might discourage administrative efforts to reduce errors. Even if the error rate is reduced as a result of such efforts, increasing administrative costs would have an opposing upward effect on liability. The cost-effectiveness of error-reduction efforts would determine whether liabilities increase or decrease. The "harder" types of errors--those whose prevention requires greater administrative expenditures--become relatively unrewarding to address, because they involve greater increases in the administrative cost base on which liabilities are calculated.

An alternative would be to base the amount of the liability on the cost of erroneously issued benefits. Under this approach, the liability assessed for a State is in direct proportion to the cost of its certification errors. The cost of food stamp certification errors would thus be directly shared between the States and the federal government.

Once the base on which liabilities are assessed is established, the next step is to define the percentage of benefit overissuance to be assessed as a liability. This "liability rate" could be 100 percent, or it could be less. In essence, the liability rate defines the percentage of overissuance costs beyond the threshold that are borne by the States and the federal government.

Table IV.7 shows how several different options for basing liabilities on overissuances would affect total liabilities for 1985, and, for comparison, again displays the liabilities computed with the same methods based on administrative costs. The figures for liabilities based on overissuance assume a 100 percent liability rate in excess of a 5 percent threshold. Under such an arrangement, the federal government, in effect, bears the full cost of overissuances up to a payment-error rate of 5 percent, and State agencies bear the full cost beyond that point. Alternative plans, with liability rates less than 100 percent, would lead to lower liabilities.

Tying liabilities to benefits rather than administrative costs can be viewed as removing a disincentive in current policy to increase spending on administrative functions that are designed to reduce errors. It can also be viewed as an approach that would give State agencies a clearer cost/benefit framework in which to make decisions on increasing administrative spending to reduce errors. Take, for example, a hypothetical State that has had recent error rates of 9 percent, total benefit issuances of about $500 million, and total administrative costs of about $40 million (with a State share of $20 million). Such a State, faced with liabilities based directly

Table of Contents

TABLE IV.7

FY 1985 LIABILITIES UNDER ALTERNATIVE METHODS BASED ON ADMINISTRATIVE COST AND BENEFIT OVERISSUANCE

State	Payment Error Rate (r)	Liabilities Based on Administrative Cost: Alternative Policies — Point Estimate	Lower Bound	Point Estimate If Lower Bound Greater than .05	Liabilities Based on Issuance Error: Alternate Policies — Point Estimate	Lower Bound	Point Estimate If Lower Bound Greater than .05
Alabama	0.1350	$ 13,118,714	$ 9,620,390	$ 13,118,714	$ 27,042,073	$ 22,070,309	$ 27,042,073
Alaska	0.1353	2,096,708	810,883	2,096,708	2,096,708	810,883	2,096,708
Arizona	0.0938	4,329,756	3,092,683	4,329,756	5,299,307	3,826,511	5,299,307
Arkansas	0.0788	1,242,979	828,653	1,242,979	3,620,775	1,345,847	3,620,775
California	0.0708	13,136,972	4,378,991	13,136,972	13,301,523	5,622,132	13,301,523
Colorado	0.0848	1,354,275	541,710	1,354,275	3,270,976	1,802,091	3,270,976
Connecticut	0.0704	1,025,885	341,962	1,025,885	1,269,887	419,965	1,269,887
Delaware	0.0717	246,819	0	0	468,285	0	0
District of Columbia	0.0981	1,561,937	669,401	1,561,937	1,918,587	1,183,700	1,918,587
Florida	0.0671	2,432,062	1,216,031	2,432,062	6,290,533	1,994,025	6,290,533
Georgia	0.1291	16,441,248	13,911,825	16,441,248	22,906,267	17,570,931	22,906,267
Guam	0.0533	27,912	0	0	60,444	0	0
Hawaii	0.0435	0	0	0	0	0	0
Idaho	0.0516	57,098	0	0	57,098	0	0
Illinois	0.0816	9,029,457	5,417,674	9,029,457	22,534,969	14,440,565	22,534,969
Indiana	0.1090	5,659,493	4,401,828	5,659,493	14,254,580	11,035,340	14,254,580
Iowa	0.0841	2,028,618	811,447	2,028,618	3,659,096	2,141,054	3,659,096
Kansas	0.0816	1,078,122	431,249	1,078,122	2,032,863	710,055	2,032,863
Kentucky	0.0600	776,939	534,763	776,939	3,320,476	534,763	3,320,476
Louisiana	0.0976	7,719,113	5,513,652	7,719,113	17,372,318	11,848,943	17,372,318
Maine	0.0791	598,696	399,131	598,696	1,800,913	854,134	1,800,913
Maryland	0.0737	2,531,992	1,687,995	2,531,992	4,061,759	2,849,487	4,061,759
Massachusetts	0.0971	5,860,198	4,185,856	5,860,198	8,144,836	5,243,303	8,144,836
Michigan	0.0735	4,563,908	3,042,605	4,563,908	12,710,419	5,770,530	12,710,419
Minnesota	0.0951	3,218,388	1,379,309	3,218,388	4,717,328	2,239,633	4,717,328
Mississippi	0.0798	1,816,892	1,211,261	1,816,892	7,860,239	3,608,061	7,860,239
Missouri	0.0523	487,902	0	0	487,902	0	0
Montana	0.0744	385,539	257,026	385,539	760,469	432,345	760,469
Nebraska	0.0904	1,152,601	493,972	1,152,601	1,781,884	1,078,106	1,781,884

Table of Contents

TABLE IV.7 (continued)

State	Payment Error Rate (r)	Liabilities Based on Administrative Cost: Alternative Policies — Point Estimate	Liabilities Based on Administrative Cost: Alternative Policies — Lower Bound	Liabilities Based on Administrative Cost: Alternative Policies — Point Estimate if Lower Bound Greater than .05	Liabilities Based on Issuance Error: Alternate Policies — Point Estimate	Liabilities Based on Issuance Error: Alternate Policies — Lower Bound	Liabilities Based on Issuance Error: Alternate Policies — Point Estimate if Lower Bound Greater than .05
Nevada	0.0248	0	0	0	0	0	0
New Hampshire	0.0442	0	0	0	0	0	0
New Jersey	0.0850	5,829,207	3,497,524	5,829,207	9,113,638	6,457,924	9,113,638
New Mexico	0.0883	1,620,452	972,271	1,620,452	3,380,870	2,480,570	3,380,870
New York	0.0711	16,280,441	5,426,814	16,280,441	19,787,484	7,137,561	19,787,484
North Carolina	0.0649	1,802,557	61,500	1,802,557	3,531,186	61,500	3,531,186
North Dakota	0.0353	0	0	0	0	0	0
Ohio	0.0743	3,690,595	2,460,396	3,690,595	16,948,776	10,408,850	16,948,776
Oklahoma	0.1058	5,312,273	2,951,263	5,312,273	7,455,006	4,619,899	7,455,006
Oregon	0.0941	3,800,149	1,628,635	3,800,149	6,251,919	3,406,800	6,251,919
Pennsylvania	0.0936	11,709,304	5,018,273	11,709,304	23,846,342	13,139,827	23,846,342
Rhode Island	0.0800	391,265	260,844	391,265	1,054,427	562,976	1,054,427
South Carolina	0.1210	8,319,451	5,759,620	8,319,451	13,784,962	10,910,506	13,784,962
South Dakota	0.0315	0	0	0	0	0	0
Tennessee	0.0639	2,058,553	761,209	2,058,553	3,898,600	761,209	3,898,600
Texas	0.1038	28,120,597	15,622,554	28,120,597	37,710,772	26,064,943	37,710,772
Utah	0.0726	583,204	126,604	583,204	897,227	126,604	897,227
Vermont	0.0806	410,263	164,105	410,263	614,394	201,535	614,394
Virgin Islands	0.0973	299,390	213,850	299,390	1,091,184	692,717	1,091,184
Virginia	0.0667	1,415,766	707,883	1,415,766	3,154,476	1,321,197	3,154,476
Washington	0.0950	4,048,211	2,891,580	4,048,211	6,312,901	4,628,268	6,312,901
West Virginia	0.0507	111,525	0	0	111,525	0	0
Wisconsin	0.0800	1,267,661	845,107	1,267,661	4,448,693	2,424,019	4,448,693
Wyoming	0.0678	138,332	0	0	264,729	0	0
Total	0.0830	$ 201,189,415	$ 114,550,326	$ 200,119,826	$ 356,761,623	$ 214,839,616	$ 355,311,640

on its erroneous benefits beyond a 5 percent threshold, could estimate the likely error-reduction effectiveness of alternative investments in an improved administrative process, and determine whether such investments are likely to be cost-effective. Measures that administrators view as potentially capable of reducing the error rate by 1 percent, for instance, would appear to be cost-effective if they cost less than $5 million in State funds.

CONCLUSIONS

- Using administrative costs as the basis for assessing liabilities creates inequities because States that spend more on administration (relative to benefits) tend to face higher liabilities relative to their total overissuances.

Table of Contents

By collecting data on critical organizational functions, a quality control system provides a basis not only for measuring performance but also for formulating management decisions to improve performance. Although the present legislative mandate for the FSP quality control system emphasizes performance measurement and accountability, part of the motivation for quality control is to provide program managers with useful information to help them improve program management and reduce the incidence of errors. Even with the current legislative emphasis on performance measurement, the system creates a wealth of information about caseloads and errors that can be tapped by program managers for their own purposes. This chapter examines the QC system from this second perspective, assessing how well it serves as a management information system for State program managers.

In the broadest sense, a management information system (MIS) should help managers make decisions to promote the goals of their organization. Although systems could be designed to provide information relevant to every goal, the quality control system, by design, deals with only one: the accurate determination of eligibility and benefit amounts.76/ Thus, in assessing the MIS functions of the QC system, FNS has focused on the capacity of the system to provide information that supports management efforts to reduce error rates as they are measured by the QC process.

QC reviews create a data base for a random sample of each State's caseload, containing two types of information:

1. Reported case characteristics and circumstances: household size, gross income by source, deductions and net income, assets, and the demographic characteristics of household members (age, race, sex, and educational level).77/

2. Error information: the dollar value of identified errors; the program "element" in error (the aspect of eligibility rules that was improperly applied); the "nature" of the

---

76/As noted in Chapter I, FNS has additional measurement and review procedures to address other key aspects of the basic objectives of the program.

77/These data are reported as contained in the case record under review.

error (the specific household circumstances or agency action which created the error); and a distinction between agency-based and client-based errors.

This information becomes available to State agencies in three stages. First, case-by-case error findings become available to the State agency when completed individual reviews are reported to the State QC unit by reviewers. Second, because the State sample of reviews accumulates monthly during the course of the year, error rates can be computed at any point during the year and at its close.78/ Third, corrections are made to individual case review findings in the resolution of differences between State review results and federal re-review results for cases in the re-review sample.79/

This chapter examines three issues associated with how this information is applied80/:

****

MANAGEMENT INFORMATION ISSUES

Issue A: How do State agencies now use QC information to support management decisions, and how do they use the QC process as a vehicle for collecting other useful information?

Issue B: What types of information would be made available to States from an "ideal" MIS?

78/Due to internal State procedures for reexamining all or some completed reviews, some delay may occur between the time a reviewer completes a specific case and the time the State makes its final determination on that case.

79/The final step in the error-rate measurement process, of course, is the calculation of the official error rate by the federal government, but this step yields only a statistical estimate pertaining to the overall food stamp caseload, not additional information on individual cases.

80/The information and analysis reported in this chapter are derived largely from Alan S. Werner, "The Quality Control System as a Management Information System for States," Abt Associates, Inc. (forthcoming).

Issue C: In what respects does the present QC system and the information it provides meet the requirements of an ideal MIS? In what ways does it fall short? How could the present QC system be extended or modified to make it fulfill MIS goals more fully?

****

A. CURRENT USES BY STATES OF THE MANAGEMENT INFORMATION GENERATED BY THE QC SYSTEM

Although the QC system has evolved in the direction of placing greater emphasis on holding States accountable for program performance, considerable effort has been devoted at the federal and State levels to making QC data and the QC data collection process a useful source of management information for State agencies. States use the QC system for management information along two dimensions: (1) using the QC data collection process to extend the information required by the standard QC review, (2) using the results of the QC data collection process to undertake analysis to support management decisions.81/

==========

Question A1: To what extent, and in what ways, do State food stamp agencies use the QC data collection process as a source of management information?

Finding: Almost half the States (25) collect additional data in their reviews beyond QC requirements, thus taking advantage of the basic sampling and review process to extend the information they can derive from it. This supplementary data collection focuses on additional demographic characteristics. In addition, 11 States report that they review supplementary samples, primarily to improve the precision of their results and to facilitate analyzing the sources of error and error rates for specific sub-State units.

Question A2: What types of analyses do States perform with QC data?

Finding: The most common analysis identifies error-prone types of households (71 percent of the States report performing this analysis routinely, and the remainder occasionally). Other common analytical uses entail identifying error-prone

---

81/Information on the current uses of the QC process is drawn from the "Food Stamp Operations Study Census of State Operations: Quality Control Systems," Abt Associates, Inc. (forthcoming).

offices, preparing descriptive statistics on the food stamp caseload, making caseload projections, and evaluating potential changes in policy.

==========

How do States Use QC Data as Management Information?

Quality control reviews provide not only a measure of error rates, but also a relatively rich source of information on the program caseload and its characteristics. Moreover, some States have built on the basic QC system to enhance how they use the data base created from it. Almost half of the State food stamp agencies have chosen to "piggy-back" the collection of supplementary data on the QC data collection process. For a relatively low marginal cost, States can collect supplementary data during the QC review process to complement the basic QC data and provide a richer base for management analysis. States do so in two ways: by collecting additional data in QC reviews beyond what is required, and by supplementing their QC samples with additional reviews. In addition, of course, States may choose approaches outside the QC system to generate management information that the QC system cannot.

Collecting Additional Data. Of the 53 State agencies, 25 collect additional data in QC reviews beyond what the standard QC process requires. Table V.1 lists the agencies that collect supplementary data and the categories of information that they add to the QC review. With only three exceptions, these agencies routinely collect this supplementary information on all QC sample cases. The supplementary data collection focuses most commonly on the demographic characteristics of the household reviewed. Another item commonly collected is information on whether or not the household is subject to monthly reporting, most likely of interest because the State wishes to examine relative error rates among monthly reporting and non-monthly reporting households.

Expanding the QC Sample. Another way that States can build on the QC system as a source of management information is to expand the sample for which they collect QC review data (as well as any supplementary data they may gather). Once they have satisfied federal QC sample requirements, State agencies may collect data on an additional sample and may elect to include the additional sample in their computation of error

Table of Contents

TABLE V.1

COLLECTION OF SUPPLEMENTARY DATA AS
PART OF QC REVIEWS

State	Are the Data Collected for All Cases?	Types of Data Collected: Demographic Characteristics	Welfare Participation History	Work Experience	Monthly Reporting	Caseworker Identification	Utility/ Heating Sources	Other[b]
Alabama	YES	X						
Arkansas	YES	X				X	X	X
California	YES	X						
District of Columbia	YES	X	X	X				
Florida	YES				X			
Indiana	YES				X			
Kentucky	YES	X		X				
Maryland	YES		X	X				X
Massachusetts	NO[a]							X
Michigan	YES	X	X					X
Minnesota	YES	X			X			X
Mississippi	YES	X				X		
Missouri	YES	X			X			
Montana	YES					X		
Nebraska	YES		X					
New Hampshire	YES							X
New Jersey	YES						X	
New York	YES	X	X	X				
Oklahoma	YES				X			
Tennessee	NO[a]	X		X				
Texas	YES				X			
Utah	YES	X	X	X				
Vermont	NO[a]						X	
Virginia	YES				X			
Wyoming	YES				X	X		

[a]Subsamples of QC sample. In Tennessee, supplementary data are collected only in January and February.

[b]"Other" includes vehicle value, rent, supervisor identification, child support, detailed shelter costs, and student information.

rates.82/ Increasing the State sample size can enhance the analytical precision of the State results and enable them to undertake more analyses of sample subgroups with some statistical confidence.

Only 11 States reported that they do in fact supplement the QC sample with additional households for which comparable data are collected, and several of these States supplement the QC sample with small or special-purpose samples. As shown in Table V.2, most States that supplement the QC sample report do so to increase the precision of their error-rate estimates and to improve their ability to analyze the sources of error and error rates by office. Of these 11 States, 8 perform additional reviews for substantial household samples beyond the required QC sample.83/ Three States supplement their samples with relatively small numbers of households. Five States include the supplemental sample data in the QC data they report to FNS and use the supplemental sample along with the required sample to calculate their error rates. Six States use the supplemental sample only for their own analysis. Two of these six, Kansas and Pennsylvania, reported supplementing their QC samples with special samples drawn from error-prone categories of households.

Generating Management Information Outside QC. Extending the QC process--by collecting additional data in QC reviews or expanding the QC sample--is not the only method whereby States gather information on their caseloads and sources of error to help them make management decisions. A total of 18 States reported conducting case reviews using different methods than those used in the QC process. As shown in Table V.3, the sample sizes and sample selection methods used by the States

---

82/However, sampling procedures must distinguish between cases selected for formal error-rate analysis and supplementary samples not included in error-rate analysis.

83/Although no exact figure for the supplemental sample was reported by New York State, it appears that the State selects cases for every local office, a practice which most likely represents a substantial addition to the required sample.

Table of Contents

TABLE V.2

STATE SUPPLEMENTATION OF QC SAMPLES

State	Number Of Supplemental Reviews	Reason For Supplementing	Reported To FNS
Arkansas	480	To allow analysis by office	No
California	1,000	To allow analysis by office, and of causes and types of error	No
Hawaii	4,800	To allow analysis by office, and of causes and types of error; to correct errors and prevent them from appearing in formal QC sample	No
Indiana	35	To allow analysis of causes and types of error	
Kansas	1,680	To identify errors not found in QC sample	No
Massachusetts	60	Special reviews for SSI Demonstration	Yes
New Hampshire	68	To increase precision of error-rate estimates, and allow analysis of causes of error and offices	Yes
New Jersey	1,182	To increase precision of error-rate estimates, and allow analysis of causes and offices	Yes
New York	Varies by office	Analysis by local office	Yes
Oregon	578	Increase precision of estimates, and allow analysis of causes and offices (FY85)	Yes
Pennsylvania	4,000	To identify errors outside QC sample	No

Table of Contents

TABLE V.3

REVIEWS UNDERTAKEN OUTSIDE THE FORMAL QC PROCESS

State	Sample Size	Review Method	Sample Selection Method
Alabama	500	File review by QC reviewer	For selected counties, all certifications
California	20	File review	For selected offices, a random sample
Delaware	5,000	File review	Error-prone case selection model
Illinois	4,450	File review	For individual offices, a random sample of last 4 months of recertifications
Maryland	23	File review	Same as QC method
Massachusetts	20,000	File review	Random sample at individual offices
Minnesota	1,320	File review	Same as QC method
Montana	3,000	File review	Targets offices with highest error rates
New Jersey	1,000	File review	Random sample of last 4 months of recertifications targeted toward 2 largest offices
New York	DK	File review	Varies by local office
North Carolina	DK	File review	Targets offices with high error rates
North Dakota	500	File review	Random sample of cases of particular types (e.g., NPA) by selected offices
South Dakota	8,000	File review	Error-prone case selection model

Table of Contents

TABLE V.3 (continued)

State	Sample Size	Review Method	Sample Selection Method
Texas	200	DK	DK
Utah	4,200	File review	Error-prone case selection model
Virginia	4,500	Informal review by policy specialist	DK
Wisconsin	300	File review	Error-prone case selection model
Wyoming	Ad Hoc	File review	Targets special problems (e.g., workers with high error rates)

DK = Don't know.

vary quite extensively.84/ Without exception, however, these States rely on reviews of case files, a more limited and less expensive approach than the full QC review, which may entail a home visit, interviews with households, and contacts with third parties.

Thus, reviews undertaken outside the standard QC process are likely to be less expensive than QC reviews, although more limited. Most of these States did not have estimates of the time devoted to these special reviews readily available. However, the 8 States that did offer such time estimates for these reviews estimated that they devoted an average of only 20 percent of the staff time normally devoted to the standard QC review. Clearly, given this cost difference, States are more likely to broaden their analysis of certification errors by reviewing large samples of households with methods that depart from the formal QC review process rather than by drastically increasing the size of the QC review sample.

QC Data for Analytical Purposes

The fairly widespread practice of supplementing the QC process and the QC sample indicates that States have an interest in using the basic QC data as a starting point for analyses to support management decisions. Table V.4 summarizes the percentage of States that reported using QC data for a variety of analyses on a routine or occasional basis or not at all. However, these data should be interpreted cautiously, because most of the respondents who were interviewed gave little indication about the degree to which any particular type of analysis is undertaken, the extent to which the analytical results are used, or the degree to which managers rely on these results as they make decisions about certification procedures or resource allocation.

However, one fairly clear pattern emerges--that analyses of QC data are most commonly used to identify error-prone types of households and of offices that exhibit high error rates. Moreover, States routinely use the error-prone profile software provided to them by FNS.

84/Some of the States reporting special reviews may have been referring to reviews that are required in Management Evaluations, which must be conducted annually in large project areas (caseloads greater than 7,000), every two years in mid-size project areas, and every three years in small project areas (caseloads less than 250). These management evaluations must include case file audits.

Table of Contents

TABLE V.4

PERCENTAGE OF STATES USING QC DATA
FOR VARIOUS ANALYTICAL PURPOSES

Analytical Purpose	Extent of Use		
	Routine	Occasional	Never
Identify Error-Prone Cases	71%	29%	0%
Identify Error-Prone Workers or Groups of Workers	20%	22%	59%
Identify Error-Prone Offices	71%	16%	14%
Describe the Food Stamp Caseload	41%	41%	18%
Project Caseload Size and Characteristics	35%	27%	37%
Evaluate Changes in Program Policy/Administration	34%	38%	27%
Project Effect of Policy and Procedural Changes	26%	36%	38%

CONCLUSIONS

- Most States now use QC for management information. The most common use is to identify error-prone cases and offices. A number of States also use QC data to describe and project caseloads and evaluate the effects of policy changes.

- Many States supplement the basic QC process--by collecting additional data or reviewing additional cases--to strengthen management information. Others extend the information available to managers by using methods other than QC. Such methods typically entail less intensive desk reviews of more cases.

B. REQUIREMENTS FOR AN OPTIMAL MANAGEMENT INFORMATION SYSTEM

Current QC practice at the State level provides some evidence that the QC review process and the data it generates are used by program managers for MIS purposes. However, given that the primary purpose of the system is to measure error rates, the degree to which this system satisfies management information needs remains an open issue. Thus, assessing the usefulness of the QC system as a management information source must begin by defining the requirements that must be met by an ideal MIS.

==========

Question B1: What are the primary requirements of an ideal QC management information system for State program managers?

Finding: The ideal MIS should (1) allow progress towards error-rate targets to be measured, (2) allow performance to be analyzed at lower (sub-State) administrative levels, and (3) help managers identify ways to improve performance.

==========

Characteristics of an Ideal MIS

To provide optimal support to program managers, a management information system should serve three broad functions. First, it should provide performance measures which facilitate measuring progress towards a performance goal. Second, the system should give program managers information which enables them to assess performance, establish accountability, and measure progress at lower administrative levels. Third, the system should generate data to help managers identify ways to improve performance.

To help managers measure progress towards performance goals, the ideal management information system should meet several requirements. The methods used to measure error rates and the underlying definitions of error must be relatively stable, so that changes in measured error rates can reasonably be viewed as changes in performance rather than changes in the measurement focus or technique. Performance measures must also be available on a schedule that corresponds approximately to the cycle on which managers can realistically expect to review performance and take new initiatives or remedial action. For most managerial decisions aimed at reducing error rates, this criterion probably implies that performance measures are needed at quarterly intervals, to allow managers some oppor-

tunity to undertake mid-year evaluations of progress towards annual error-rate targets.

The ideal management information system would also provide performance measures for sub-State administrative units, so that managers could identify the relative success of particular components of their agency at meeting error-rate targets. This information would enable them to focus their corrective action efforts and resources--for example, management attention, additional staff, extra training--on particular components of their agencies.

To help program managers identify ways to improve performance, the ideal management information system should provide information that enables them to make decisions on allocating resources and developing or modifying agency procedures or operations. Allocating resources may entail focusing existing procedures or staff attention on certain types of cases, or shifting the frequency or intensity of certain procedures (e.g., verification, monthly reporting, home visits, and computer matching).

C. ADEQUACY OF QC AS A SOURCE OF MANAGEMENT INFORMATION

In assessing the QC system as an MIS for States, FNS examined how well it fulfills the requirements of the optimal MIS described in Section B. In this assessment, FNS focused on how well the QC process provides interim measures of progress, provides performance measures for sub-State administrative units, and helps identify ways to improve performance. For each of these three requirements, FNS examined how the adequacy of the QC system as an MIS is affected by the timing of QC results, the size of review samples, the content of the data collected, and the analytical methods used on QC data.

==========

Question C1: How well does the QC system provide interim measures of progress toward reducing error rates?

Finding: Although final official error rates are not published until months after the end of each review year, States can compile results from reported reviews during the course of each year. Few States have large enough samples to make monthly results meaningful, but 39 States have large enough samples to make reasonably satisfactory statewide estimates of error rates on a quarterly basis.

Question C2: How well does the QC system provide error-rate measures for sub-State administrative units?

Finding: Current sample sizes severely restrict the precision attainable in sub-State error-rate estimates; only a handful of local offices nationwide could be expected to be represented in State QC samples by enough cases to obtain any reasonable precision for an office-specific error-rate estimate.

Question C3: How well does the QC system help identify ways to improve performance?

Finding: The QC system helps identify the types of errors most commonly committed and the types of households most susceptible to error. However, it does not collect sufficient data to identify the

types of household actions or agency procedures that cause error, or the types of procedures that could help prevent errors.

==========

Interim Measures of Progress towards QC Goals

The QC system has at least the potential for providing interim measures of error rates and of progress toward reducing errors during the course of each annual review cycle. Although final error rates based on the federal re-review become available nine months after the end of the review year (beginning with 1986 rates), State review results are compiled each month during the review year, and can be analyzed by State agencies periodically during the year to generate interim measures. State review results are of course modified by the federal re-review, but not to degrees which would seriously affect answers to the basic questions that are of concern to State program managers during the course of a year: Is the State's interim error rate high or low? Is it seriously above the annual target rate? How does the interim rate compare with last year's annual reported error rate and the corresponding interim rate from the year before?

Without waiting for the final results of the federal re-review or the final official error rate, State agencies could generate results on their own early enough to provide interim measures. High error rates detected by interim measures are a sufficient basis for management action, even in the absence of final results pertaining to fiscal liabilities. Moreover, as federal re-review cases are completed during the year, States could even compute an "interim regressed error rate," applying the same formulas used by FNS.

An important issue, however, is whether review sample sizes are adequate to yield interim measures of sufficient precision to be useful to program managers. Suppose, for instance, that the requirement is established that interim measures must have "relative confidence intervals" no greater than 50 percent of the estimated error rate.85/ In general, meeting this standard requires a sample size of about 200 cases. Only 11 States currently have large enough QC samples to yield a monthly interim measure that meets this standard. However, quarterly error-rate estimates would probably provide frequent enough

---

85/This requirement would mean that if, for example, the estimated error rate were 12 percent, the confidence interval should be no larger than plus or minus 6 percent.

readings for most program managers, and the samples of 39 States are large enough to meet this standard.

Alternative strategies could be used by States to improve their ability to produce interim measures despite sample size issues. For example, States could use rolling samples, with each interim measure based on the past six months of review (perhaps even extending across review years). States could also use a cumulative sample, with each interim measure based on all months to date in that year, which would produce narrower confidence intervals throughout the course of the year.

From the standpoint of data content and analytical methods, interim measures are certainly feasible. As long as the accepted overall performance measure is the QC error rate, the system yields the necessary data, and the normal analytical procedures used to create the reported error rate can be applied at intervals during the year to generate interim measures.

Performance Measures for Sub-State Administrative Units

From the standpoint of data content, timing, and analytical results, the QC review process can generate performance measures for administrative units within a State just as it can for the entire State. As with interim measures, however, the size of regular State review samples poses the primary obstacle to using the present QC process as a basis for developing sub-State estimates with reasonable statistical precision.

If State managers wish to estimate error rates for individual local offices, they can expect reasonable precision only in rare instances, given the small sample sizes that are drawn from individual offices. If the standard described earlier is used (a requisite 200 review cases to yield a relative confidence interval of 50 percent), acceptable results would be derived only from a handful of offices. As illustrated in Table V.5 for eight selected States, office-level estimates that meet this standard of statistical precision cannot be expected in more than one or two offices at most, and in many States no such estimates could be obtained.

Expanding the QC sample can ameliorate this sample size constraint to some degree, but only large sample increases could be expected to allow States to generate precise estimates for many local offices. An alternative available to States, of course, is to conduct more limited reviews--based only on the case file review, without a field review--for a substantial sample of cases.

Table of Contents

TABLE V.5

ANNUAL QC SAMPLES SIZES FOR LOCAL OFFICES
IN SELECTED STATES

State	Number of Local Offices	Offices with > 200 Cases	Offices with 100-200 Cases
Alabama	67	2	1
California	53	1	4
Delaware	22	0	0
Michigan	113	0	1
Nevada	14	1	0
New York	150	1	0
Utah	45	0	0
Wyoming	22	0	0

Identifying Ways to Improve Performance

The challenge facing State program managers in using QC data is to analyze the data in ways that shed light on the management issues they face and help them allocate program resources to correct problems. Ideally, an MIS would help program managers answer five questions about their operations:

1. What types of errors occur most often, and what types of household information are most often missing or incorrect?

2. What types of households are most likely to be in error?

3. What management or administrative characteristics within the State are associated with higher rates of error?

4. What household actions or agency procedures appear to cause errors or allow errors to go undetected?

5. What agency procedures could have been introduced to prevent or correct observed errors?

The present QC system serves management information needs quite well for questions 1 and 2 but provides very limited information to respond to questions 3 through 5.

The present QC system can make a very limited contribution to identifying ways to improve performance. QC reviews themselves do not capture data on the management or administrative setting in local offices within a State. Moreover, attempting to use measures of workload, skill levels, or organization to explain the incidence of error raises very difficult methodological problems. It is simply difficult to define and measure appropriate dimensions of the local work environment.

Similarly, the QC system lacks data on household actions or agency procedures that cause errors. When errors are observed, the detailed coding of the "nature" of the error sometimes identifies a particular procedural failure that contributed to the error. However, information on whether or not a procedure should have been applied to a case, and whether it was actually performed, is absent from the review results for correct cases and, in many instances, for error cases. Failure to perform a particular procedure (e.g., verification) may not cause an error and would never be recorded in the QC review, even when an error arises due to some other problem.

The present QC system cannot directly contribute significantly to managers' understanding of the procedures that could have prevented observed errors. Although the review schedule records information on how the reviewer discovered the error, the codes for recording this information are limited. And the reviewer's discovery procedure might not have been available to the agency. The present system does not gather any information on the reviewer's judgment about what agency procedures might have prevented the error, or what might have prevented the procedural failure from occurring.

Despite these shortcomings, the present QC system can help program managers analyze two of the important questions identified above that can contribute to improved performance: (1) what types of errors occur, and (2) what types of cases are in error? The manner in which the QC system provides information on these issues, and the analytical methods that can be applied to QC data to answer these questions as fully as possible, are examined below.

Analyzing Types of Errors. Program managers who seek to improve performance must understand which eligibility rules have been violated, and in what way the errors were committed--for example, through arithmetical error, failure to apply rules, or the misapplication of a particular rule. Ideally, two types of information are necessary to clarify these issues. First, the manager would like to know what proportion of errors are attributable to any given type of error, in terms of both their number and cost. Just as important, however, is information on the frequency of correct or incorrect determinations given the applicability of a specific program rule. A particular program element may be a frequent source of error either because it causes problems in a high percentage of circumstances in which it is relevant to the certification or because it is frequently relevant. Distinguishing between these two contributions to the frequency of errors can help managers decide what type of corrective action should be initiated.

Analyzing of the sources of error--focusing on which elements are the most frequent sources of error--is generally feasible given present QC review data. Analysis should be able to distinguish which of approximately 40 program elements exhibit particularly high or low error rates. For example, to detect whether the error rate for a particular element is half a percentage point above the State's overall error rate for all elements at the 95 percent confidence level requires a sample size in the range of 900 to 1,100 review cases, and at the 90 percent confidence level in the range of 475 to 600 cases. These requirements can be met in many States: 38 States have annual samples of 1,000 or more, and 44 have annual samples of 500 or more.

QC data can also be used to generate information on the second issue described above: the frequency with which cases are handled correctly, given that a particular program element applies. Although the QC system does not systematically record the applicability of particular elements when they have not given rise to an error, a detailed examination of the QC review schedule shows that, in most cases, the applicability of particular elements can be inferred from the data on case characteristics. "Correct handling rates" can thus be computed for about 30 of the 40 program elements in the review schedule.

Current QC sample sizes can often support reasonably precise estimates of correct handling rates. The adequacy of sample sizes varies not only across States, but also across program elements, since some elements apply to large segments of the caseload and others to very small segments, and it is the number of cases for which the element is relevant that determines the effective available sample for which the handling rate is determined. Each State's QC sample is thus adequate for a precise analysis of handling rates for some elements but not for others. For detecting element-specific correct handling rates at half a percentage point above or below the State's average rate for all elements, current QC samples are adequate to analyze 35 to 50 percent of program elements, depending on the overall sample size of the State. No State presently analyzes correct handling rates, but such analysis could be a useful source of additional information from the current QC system.

Analyzing the Case Characteristics That Contribute to Errors. Identifying the characteristics of households most prone to error can provide a basis for program managers to target resources by applying more intensive certification procedures, assigning experienced eligibility staff, and designing special interviewing or investigative methods. Two steps are involved in using QC data in this way. First, QC results must be analyzed to determine which characteristics are associated most frequently with errors. Second, the results of this analysis must be reduced to readily applied categories of households, or key characteristics, which can be used during the intake or recertification process to trigger special procedures. In addition, of course, program managers can structure staff training curricula to focus on these

special procedures.86/ State agencies commonly use QC data in this way, applying methods known broadly as "error-prone profiling." In the Program Operations Study, 37 States reported using error-prone profiling routinely, and the rest reported conducting such analysis occasionally.

However, a complete assessment of the value of QC results for this MIS function should examine the statistical precision attainable in this type of analysis. Some standard of acceptable precision must first be defined. For purposes of this discussion, FNS took as a standard the ability to detect error-rate differences between defined case categories equal to one-half of that category's overall error rate. For example, if the overall error rate is 10 percent, we would want to be able to detect with confidence error-rate differences of 5 percentage points.

The adequacy of State QC samples for detecting such differences depends on the desired confidence level, the desired power (that is, the probability of detecting a true difference of the specified size), and the variance of the dependent variable (payment error dollars). Based on an analysis performed for FNS, the annual QC samples of most States are large enough to yield reasonable precision in analyses of the relationship between specified case characteristics and error rates. However, some small States, and States that exhibit very low error rates, would have to combine two years of data to generate results that meet the defined standard of precision. However, the precision of such analysis varies considerably. States should thus exercise some caution in interpreting apparent relationships between case characteristics and error rates and making management decisions based on such analysis.

86/One limitation on the impact of these steps is obvious--that households which do not report information (e.g., the presence of income) that would reveal them as meriting special attention cannot be identified for application of the special error-prevention measures.

CONCLUSIONS

- The current QC sample sizes of most States are large enough to generate quarterly measures of interim progress, but are less useful for measures of the performance of local offices. All States can increase their sample size to improve the reliability of interim or sub-State measures, and the federal government will pay half of the costs.

- QC data are reasonably complete with respect to describing the characteristics of the caseload and the types of errors that occur, but more limited with respect to the management characteristics or procedures that cause or prevent errors.

Table of Contents

The preceding chapters explored particular aspects of the food stamp quality control system in great detail. It is clear that numerous interdependent technical and policy choices have been made in both constructing and using the quality control system for specific policy purposes. This study has examined two major ways in which QC can improve certification accuracy: (1) as a performance measurement system that provides a reasonable basis for holding States accountable for errors, and (2) as a source of management information that can help program managers identify sources of error and take steps to remedy them. The report presents the results of numerous special studies and analyses of the adequacy of the QC system in terms of these two major functions. This chapter summarizes the conclusions reached by USDA.

A. CONCLUSIONS ABOUT THE QUALITY OF THE STATISTICAL DESIGN AND THE DATA

Chapter III examined the quality and adequacy of the data that form the basis of the quality control measurement system. The basic question is whether the system is designed with a technically sound measure of performance at its foundation. Given that, the next question is whether the system is operated in a manner whereby it produces accurate and reliable results. The broad answer to both questions is yes.

The statistical foundations of the QC process are sound. The sampling process is a standard application of widely accepted methods and is carried out with a generally high level of accuracy, reliability, and consistency. The statistical procedures used to estimate official error rates are unbiased and do not systematically overstate or understate error rates. Moreover, using results from both the State reviews and the federal re-reviews substantially improves the precision of the estimated error rates.

The sampling variability in estimates of State payment-error rates is not trivial. On average, the sampling and estimation procedures yield very accurate estimates of true error rates based on repeated samples. Of course, States select only a single sample each year. Given the sample sizes now used, a chance exists that in any one year the estimated error rate will be substantially higher or lower than the true error rate. Minor improvements in the precision of estimates can be made in a variety of ways, but large reductions in sampling

error can be achieved only by substantial increases in sample sizes.

In sum, the system produces a technically sound measure of error. But which components of program accuracy should be included in the definition of error? More simply, what should States be accountable for? The answers to these questions are critical: they define the goal of any accountability system. They also help ensure fairness and equity. Several conclusions emerge from the study of this aspect of the system.

The present definition of QC errors is generally satisfactory, although some modifications could create a more balanced set of incentives for reducing errors. Food stamp QC focuses on substantive errors that are associated with true costs in misspent benefits and excludes strictly procedural errors. QC also includes both agency-caused and client-caused errors because States can in fact take a variety of actions to reduce the incidence of errors due to client misreporting. Current procedures also provide an accurate measure of losses to the Food Stamp Program based on the total income and resources actually available to the household.

The error-rate definition could be broadened to include not only overpayments but also underissuances. Such a realignment of the present focus on overpayments in the official error rate would clearly extend the scope of the system to make States accountable for certification accuracy in general. While such an extension provides a clear, understandable message of the importance of overall certification accuracy, there is no empirical evidence to suggest that State efforts to reduce overpayment errors are associated with increases in underissuances. Incorporating negative action errors--improper denials or terminations of benefits--into the error-rate measure is not feasible under the current design of the system. Because negative action reviews are based on a different sampling universe--actions rather than active households--combining the results of the two types of reviews is not practical. Moreover, the negative action review is designed to identify procedural errors rather than households that were eligible but denied benefits. Major changes to the sampling and review process would be required to incorporate negative action errors into the official error rate. This, of course, does not preclude using these measures in a two-tier accountability system (indeed, this practice is currently used).

Some characteristics of caseload composition and local setting vary across States and help explain some of the differences in error rates. Such variance implies that a measure of relative administrative performance could potentially be developed. The interplay between these factors and error rates, however, is very complicated. An extensive analysis found no clear empirical basis for adjusting State error rates on the basis of these factors. The analysis showed, on the one hand, that certain caseload and local area characteristics (household size, the presence and source of income and assets, the number of deductions, and population density) have significant effects on components of the payment-error rate.

On the other hand, the statistical models can be specified in a variety of equally plausible and useful ways that affect different States in different ways. The choice of a particular adjustment model, therefore, implies a particular set of financial consequences, yet no clear empirical basis exists for selecting one model over another. Any selective focus on a limited set of factors runs a sizable risk of introducing inaccuracy into the QC measurement process. Adjusted error rates in such a system would not necessarily be better, only different.

B. CONCLUSIONS ABOUT THRESHOLDS AND LIABILITIES

Given a sound approach to measuring State error rates, how is the cost of certification error allocated appropriately between the State and federal governments? Three major policy issues are examined in Chapter IV: (1) where the error-rate threshold should be set, (2) how the comparison between estimated error rates and this threshold should be made, and (3) how the amount of liability for a State whose error rate is above the threshold should be computed.

The basic idea of setting a liability threshold, and thus of allocating some of the cost of certification errors to State agencies, is basically sound. Given the large costs of error--more than $900 million in 1985--it is appropriate to continue a policy under which States pay at least some portion of the cost of errors. Setting the level of the liability threshold, however, is essentially a policy judgment; a "correct" threshold cannot be determined empirically. With the current threshold of 5 percent (and current methods for computing the amount of the liability), State agencies are accountable for less than 25 percent of the total cost of erroneous issuances.

Although no empirical basis exists for setting a particular threshold, there is evidence that thresholds need not be the same in the Food Stamp and AFDC programs. Differences in caseload characteristics explain virtually all of the differences between food stamp error rates and the historically lower observed rates in AFDC.

This study also paid close attention to the method for comparing error rates with the liability threshold to determine whether States should be subject to a liability and how the liability amount should be computed. For States whose true error rates are just under the threshold, sampling error creates a substantial possibility that the estimated error rate will fall above the threshold and a liability assessed. And for States whose true error rates are just over the threshold, there is a substantial chance that the State will avoid a liability that should be assessed. Because States receive no "credit" for error rates below the threshold, average liabilities are overstated relative to the liability that would be determined if the true error rate were known.

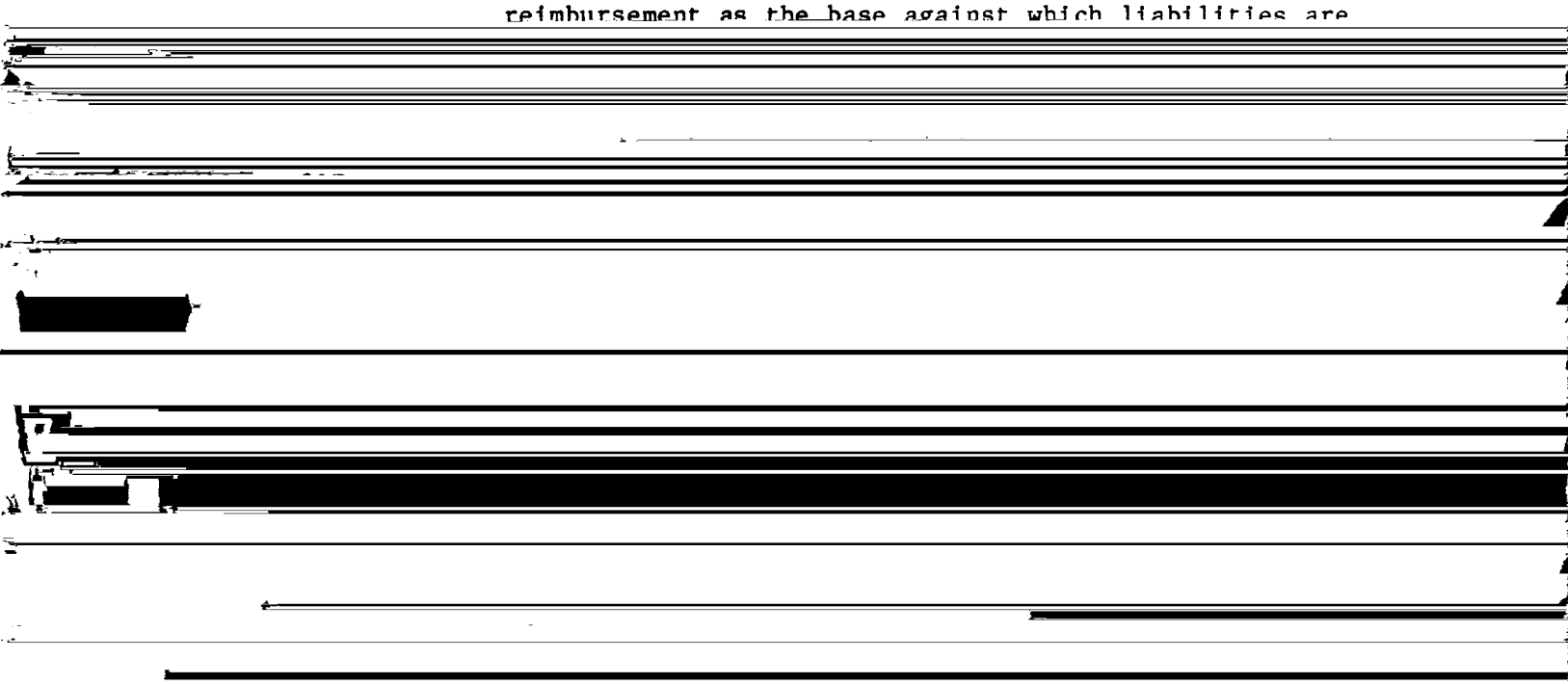
Several options would protect States and the federal government to varying extents from the consequences of sampling error. The lower bound of the confidence interval, rather than the point estimate, could be compared with the threshold, but this approach systematically understates true error rates. Another alternative would be to continue to use the point estimate, but give States "credit" for error rates below the threshold, in the form of "negative liabilities" that could be subtracted in later years from liabilities assessed for error rates over the threshold.

The analysis described in Chapter IV identified two disadvantages of present legislative policy for establishing liability amounts. First, it was shown that the "step function" for computing liability amounts systematically overstates the true liability for some States and understates it for others. Moreover, because liabilities increase in increments of 5 and 10 percent of federal administrative reimbursement, States which exhibit very similar estimated error rates--and possibly even the same true error rate--can be subject to sharply different liability proportions. This feature raises important questions about the computation method.

Both concerns--the overstatement or understatement of the true liability amount and the differential treatment of States in comparable circumstances--could be addressed by replacing the step function with a smooth function. Under this alternative, States would still be subject to liabilities if the point

estimate of the error rate were above 5 percent, but the amount of the liability would be based on the exact error-rate estimate rather than on the interval in which it fell.

The second aspect of liability computation that was examined closely in Chapter IV is the use of administrative cost reimbursement as the base against which liabilities are



program administration in an effort to reduce errors; it is possible that additional efforts to reduce errors increase liabilities.

In sum, a smooth function, if coupled with symmetric liabilities and credits based on the point estimate, would produce liabilities that, an average, nearly equal the liability that would be assessed if the true error rate were known. In other words, this approach would distribute the risk of sampling error equally between States and the federal government. Basing the liability amount on the amount of benefits issued in error would establish a closer relationship between the true cost of errors and the fiscal consequences of errors for State agencies.

C. CONCLUSIONS ABOUT MANAGEMENT INFORMATION

The QC system clearly provides useful management information. States can and do use QC data to identify error-prone types of households. States use QC data to obtain a more in-depth understanding of the types of households served and their needs, and to identify problems with service. This type of analysis is usually carried out on a State and local basis. Some States supplement their QC samples to increase the precision of the error-rate estimates they can derive for certain types of cases or individual offices in their State.

D. RECOMMENDATIONS

In sum, the results affirm confidence in the technical aspects of QC. Many questions, however, do not have right or wrong answers. For some of these questions, the weight of the evidence enables USDA to reach clear conclusions. For others, the appropriate response is less obvious. Because the National Academy of Sciences is likely to address some of the same issues in its parallel report, USDA has not made specific recommendations for change. Instead, the Department will continue its deliberations, considering the results of this study as well as those of the NAS study to identify the most appropriate ways to operate quality control.